Printed in U.S.A.
Vol. 111, No. 2

DATA-GENERATING PATENTS

Brenda M. Simon & Ted Sichelman

ABSTRACT—Patents and trade secrets are often considered economic substitutes. Under this view, inventors can decide either to maintain an invention as a trade secret or to seek a patent and disclose to the public the details of the invention. However, a handful of scholars have recognized that because the patent disclosure requirements are not always rigorous, inventors may sometimes be able to keep certain aspects of an invention secret, yet still receive a patent to the invention as a whole. Here, we provide further insight into how trade secrets and patents may act as complements. Specifically, we introduce the concept of "data-generating patents," which refer to patents on inventions involving technologies that by design generate valuable data through their operation or use. For instance, genetic tests and medical devices produce data about patients. Internet search engines and social networking websites generate data about the interests of consumers. When data-generating inventions are patented, and the patentee enjoys market power over the invention, by implication, the patentee also effectively enjoys market power over the data generated by the invention. Trade secret law further protects the patentee's market power over the data, even where that data is in a market distinct from the patented invention and especially after the patent expires or is invalidated. We contend that the use of patents and trade secrets as complements in this manner may sometimes yield socially harmful results. We identify the conditions under which such results occur and make several recommendations to mitigate their effects.

AUTHORS—Brenda M. Simon, Associate Professor, Thomas Jefferson School of Law; Non-Resident Fellow, Stanford Law School. Ted Sichelman, Professor, University of San Diego School of Law. We thank Dan Burk, Colleen Chien, William Hubbard, Dmitry Karshtedt, Anna Laakman, Bert Lazerow, Mark Lemley, Orly Lobel, Jonathan Masur, Lisa Ouellette, Nicholson Price, Arti Rai, Lisa Ramsey, Sandra Rierson, Rachel Sachs, Josh Sarnoff, Andrew Schwartz, David Schwartz, Steve Semeraro, Jacob Sherkow, Kathy Strandburg, and Marketa Trimble, as well as participants at the IP Scholars 2014 and 2015 conferences and the University of San Diego 2015 IP Speaker Series, for their helpful comments and discussion.

INTRODUCTION

Advances in the aggregation and analysis of large data sets—"big data"—are providing a new source of value for many patent holders. For example, in the genetic testing space, Myriad Genetics leverages its diagnostic patents for breast and ovarian cancer to generate large, proprietary databases of patient information.[1] As the main provider of these genetic testing services, Myriad has amassed detailed clinical data that it maintains as a trade secret.[2] Even though Myriad's patents have effectively been invalidated, Myriad has gained a strong market advantage through the ability to mine this private database of information that its competitors cannot.[3]

In the information technology realm, Google leverages data generated by its patented search technology to afford it a competitive advantage.[4] Specifically, Google utilizes a wealth of information about its users, such as previous search queries, locations, social networking data, and other personal information, which it may maintain as a trade secret well beyond the expiration date of its patents, to improve its current search queries and its targeted advertising to those users.[5]

In this Article, we introduce the concept of "data-generating patents" to describe patents over inventions that generate unique data from users.

[1] *See infra* notes 96–115 and accompanying text.

[2] *See infra* notes 96–115 and accompanying text.

[3] *See infra* notes 96–115 and accompanying text.

[4] *See, e.g.*, U.S. Patent No. 6,285,999 (filed Jan. 9, 1998) (issued Sept. 4, 2001) (disclosing the method for Google's "PageRank" search technology).

[5] *Privacy Policy*, GOOGLE, http://static.googleusercontent.com/media/www.google.com/en//intl/en/policies/privacy/google_privacy_policy_en.pdf [https://perma.cc/U556-BV2C] (last modified Aug. 19, 2015).

Specifically, these inventions typically generate data that is distinct from the operation and use of the invention itself. To be certain, the ordinary use of a patented invention will often generate data about the invention that can be used to improve the invention itself. For instance, the use of a wheelbarrow may lead customers to provide feedback that lead to improvements in its design. Indeed, such a feedback mechanism often lies at the heart of technological progress.[6] In contrast, data-generating inventions by their operation and use may generate large amounts of data beyond the invention itself—for instance, data about users, other persons, or even the world in general—that can then be used to improve the operation of the invention or employed in a field entirely distinct from the invention.[7]

Unlike information about the invention itself—which is often disclosed in patented improvements on the original invention—data-generating inventions tend to produce data that can be maintained as a trade secret. Patent holders enjoy an increased ability to aggregate and analyze "big data" obtained through leveraging data-generating patents, and they can protect the results using trade secret protection. This presents unique legal and economic consequences that we contend may be socially problematic under certain conditions. In addition to enjoying the potentially indefinite scope of protection afforded by trade secret law for the generated data, inventors of data-generating patents need not be concerned with the risk of ordinary defenses to trade secret infringement, such as independent discovery or reverse engineering, while the patent on the underlying invention is in effect.[8] Even after the patent term ends, the data-generating patent holder may continue to benefit from the de jure lead time advantage secured by the prior patent in its compilation of data.[9]

Myriad's data-generating patents over breast cancer diagnostic tests and Google's search engine patents illustrate these concerns. The exclusive lead time afforded by patent protection has given Myriad a competitive

---

[6] *See* Karen E. Lee, Note, *Cooperative Standard-Setting: The Road to Compatibility or Deadlock? The NAFTA's Transformation of the Telecommunications Industry*, 48 FED. COMM. L.J. 487, 490 (1996) ("Manufacturers use consumer feedback to improve upon a technology.").

[7] Of course, all inventions in some rough sense "generate . . . data beyond the invention itself." For instance, the sale of a patented pharmaceutical may generate a list of physicians who regularly prescribe the drug and a list of patients who regularly take the drug. However, unlike data-generating inventions, such data is not generated, by design, through the operation and use of the invention. As such, data-generating inventions are a unique subset of inventions. *See infra* Part II.

[8] *See* Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 489–90 (1974).

[9] *Cf.* Cole M. Fauver, Comment, *Compulsory Patent Licensing in the United States: An Idea Whose Time Has Come*, 8 NW. J. INT'L L. & BUS. 666, 676 (1988) (concluding that the initial patent holder has an "advantage over competitors having to start 'from scratch' to develop the product").

advantage in generating a massive database of patient information protected by trade secrecy. Myriad's database will be extremely difficult for competitors to recreate so as to compete effectively, despite the Supreme Court's recent invalidation of many of Myriad's patent claims.[10] The inability for the market to self-correct this informational imbalance allows Myriad to extend its market power even after patent protection ends.[11] Similarly, the evidence suggests that Google has been able to leverage its patented search technology to obtain a competitive advantage in gathering data.[12] Google's ability to capture information about its users, and to continue using this proprietary information after its patents expire or are invalidated, provides it the ability not only to improve its search functionality, but also to leverage its proprietary data to create a superior product in the distinct market of targeted online advertising.[13]

Our discussion of the potentially problematic uses of data-generating patents is subject to two important qualifications. First, we do not doubt that other aspects and practices of technology companies—such as technological "lock-in," tying, collusion, resale price maintenance, reverse payments, and network effects—can lead to similar anticompetitive effects when a company enjoys market power.[14] These concerns are typically considered by antitrust law.[15] However, ordinary patent enforcement—that is, enforcement absent these other sorts of market practices—is often viewed as an "exception" to traditional antitrust scrutiny, or at least typically outside of antitrust's regulatory ambit.[16] In this regard, patent law generally uses internal doctrines to determine the appropriate reach of a patent, and although we address antitrust law when appropriate, we tend to

---

[10] *See infra* Section II.A.1.

[11] *See infra* Section II.A.1.

[12] *See infra* Section II.A.2.

[13] *Terms of Service*, GOOGLE, https://www.google.com/apps/intl/en/terms/user_terms.html [https://perma.cc/9GQN-7AU3]; *see* Nathan Newman, *Search, Antitrust, and the Economics of the Control of User Data*, 31 YALE J. ON REG. 401, 407 (2014) (explaining that Google is able to "extract ever more precise information about users to allow advertisers to more effectively target particular ads to those users"); James Grimmelmann, *Speech Engines*, 98 MINN. L. REV. 868, 942–43 (2014).

[14] *See generally* HERBERT HOVENKAMP, MARK LEMLEY & MARK JANIS, INTELLECTUAL PROPERTY AND ANTITRUST: AN ANALYSIS OF ANTITRUST PRINCIPLES APPLIED TO INTELLECTUAL PROPERTY LAW (2009); Daniel F. Spulber & Christopher S. Yoo, *Access to Networks: Economic and Constitutional Connections*, 88 CORNELL L. REV. 885 (2003).

[15] *See* HOVENKAMP, LEMLEY & JANIS, *supra* note 14; Spulber & Yoo, *supra* note 14, at 885.

[16] *See* William F. Baxter, *Legal Restrictions on Exploitation of the Patent Monopoly: An Economic Analysis*, 76 YALE L.J. 267, 312 (1966) ("It is convenient to describe the patentee's freedom to monopolize as an 'exception' to the antitrust laws' general mandate of competitive behavior . . . ."); *see also* Maureen A. O'Rourke, *Toward a Doctrine of Fair Use in Patent Law*, 100 COLUM. L. REV. 1177, 1215–16 (2000).

employ a patent policy-centric lens in evaluating the concerns of data-generating patents.[17]

Second, companies often use a variety of means to aggregate proprietary data.[18] For instance, it is clear that Google's patents today are not the predominant mechanism by which Google maintains a competitive advantage in generating its user search queries.[19] Nonetheless, as we explain below, we believe that patents did play an important role for Google early in its history.[20] Importantly, as long as data-generating patents exhibit some nontrivial detrimental effect on the market, they deserve examination and potential remediation, even if they play a relatively minor role in a given company's mix of tools and practices used to gain a market advantage.[21]

Not all data-generating patents pose policy concerns. Yet, identifying ex ante which data-generating patents are likely to create high potential social costs will typically be too difficult. So we offer two criteria that can be used to make such a determination ex post: (1) the extent of expansion into unforeseeable data markets and (2) the strength of preempting potential competition in markets for the generated data.[22]

The "unforeseeable data markets" factor concerns the patent holder's ability to generate data in an area that is not directly related to the market covered by the patented invention.[23] If the invention allows aggregation of

[17] *See* O'Rourke, *supra* note 16, at 1196. One exception is the so-called patent misuse doctrine, which tends to draw on antitrust principles to render a patent unenforceable when a patent holder engages in certain anticompetitive behaviors. *See id.* at 1195. We address potential misuse concerns of data-generating patents in Sections II.B.3 and III.B.1.

[18] *See* Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Searc*h, 93 CORNELL L. REV. 1149, 1179–86 (2008).

[19] *See* J. Gregory Sidak, *A Consumer-Welfare Approach to Network Neutrality Regulation of the Internet*, 2 J. COMPETITION L. & ECON. 349, 454 (2006); *infra* Section II.A.

[20] *See infra* Section II.A.

[21] *Cf.* Oliver E. Williamson, *The Economics of Antitrust: Transaction Cost Considerations*, 122 U. PA. L. REV. 1439, 1492 (1974) (noting that a particular market practice "has genuine economic significance, as contrasted with transitory business significance, *only* to the extent that nontrivial barriers to entry into the industry in question can be said to exist" (emphasis in original)).

[22] As we discuss below, although there are other factors that may be useful, we believe these two factors do the majority of the work in determining whether a given data-generating patent is likely to be problematic. *See infra* Sections III.A.1–4.

[23] *See* Pamela Samuelson, *The Quest for a Sound Conception of Copyright's Derivative Work Right*, 101 GEO. L.J. 1505, 1527–29 (2013) ("Unless carefully cabined to the kinds of foreseeable markets exemplified by the definitional derivatives, [copyright] can unduly restrain competition and follow-on innovation, as well as interfere with free-expression interests of subsequent creators."); Dmitry Karshtedt, *The Completeness Requirement in Patent Law*, 56 B.C. L. REV. 949, 995–96, 1004–07 (2015) (contending that patent "[c]laims to general concepts such as the hedging of risk, unconstrained by any methods of implementation, are problematic for reasons similar to functional

data in unforeseeable secondary markets, as we explain further in the Article, the economic effects of the patent will more likely be problematic.[24] For instance, Google's use of its aggregated search data derived from its patented search algorithms is potentially more problematic in our view when used to target customized advertising to users than when used to improve the algorithms per se.[25]

The preemption factor relates to the magnitude of the preemptive effect of the data-generating patent on competition in the market regarding the data. The more likely the data-generating patent forecloses competitors from generating similar data, the more likely the invention will be found problematic. For example, Myriad's data-generating patents foreclosed other methods of identifying mutations in the BRCA1 and BRCA2 genes for breast and ovarian cancer screening.[26] Economically detrimental effects from these patents are likely to present a greater concern on balance.

For those data-generating patents that are likely problematic, we discuss several proposals to mitigate their detrimental effects when they occur.[27] Because identifying problematic patents accurately ex ante will likely be difficult, we tend to prefer ex post solutions as a general matter. For the ex ante solutions we discuss—which include limiting patent term and narrowing patentable subject matter—we are concerned not only about error costs but also about an increased preference for trade secrecy that may result. Similarly, requiring disclosure of data might raise fewer of these concerns, though the effect on innovation incentives and monitoring challenges make such solutions less than ideal. For the ex post proposals, such as broadening exceptions for reverse engineering, independent invention, and experimental use, and perhaps limiting injunctive relief, we are more sanguine if courts could narrowly fashion these remedies. Similarly, agencies could also institute disclosure-related solutions. Each potential solution has possibly significant limitations. We believe, however, that any solution would benefit from detailed empirical study and further discussion.

In addition to our policy-driven analysis, we also make an original observation regarding the theoretical role patents play relative to trade secrets. Patent and trade secret law share the same goal of promoting

---

biotechnology claims: they cover a large number of avenues of further development, including some that might be unforeseeable and quite transformative").

[24] *See infra* Section III.A.

[25] *See infra* Section III.A.

[26] *See infra* Section III.A.

[27] *See infra* Section III.B.

innovation.[28] Ordinarily, patents and trade secrets are considered economic substitutes.[29] In other words, inventors decide between patent and trade secret options in a way that maximizes protection for the invention.[30] In some circumstances, however, no choice needs to be made—trade secrets and patents function as economic complements.[31] In particular, limitations in the disclosure requirements of patent law allow inventors to keep some features of their inventions secret.[32] Moreover, inventors are under no duty to update their disclosures after they file for a patent application.[33] Besides the important policy questions posed by data-generating patents, they are a unique example of patents and trade secrets as economic complements that has not been previously identified in the scholarly literature.[34] Unlike prior discussions of the complementary roles patents and trade secrets may play, data-generating patents can produce trade secrets in distinct product markets and even after the patent expires or is invalidated.[35]

In sum, our Article makes three important contributions to the literature. First, we provide strong evidence for an important class of inventions to rebut the standard view that patents and trade secrets merely act as economic substitutes.[36] Second, we identify, explain, and coin the notion of data-generating patents, which had only been described incompletely in the previous literature.[37] Third, we describe the conditions

---

[28] *See* Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 484–85 (1974); Peter S. Menell, *Intellectual Property and the Property Rights Movement*, REG., Fall 2007, at 36, 38.

[29] *See, e.g.*, Ted Sichelman & Stuart J.H. Graham, *Patenting by Entrepreneurs: An Empirical Study*, 17 MICH. TELECOMM. & TECH. L. REV. 111, 136 (2010) (explaining the traditional view).

[30] *See id.*

[31] Christianson v. Colt Indus. Operating Corp., 822 F.2d 1544, 1562–63 (Fed. Cir. 1987), *vacated*, 486 U.S. 800 (1988); *see* Gideon Parchomovsky & Peter Siegelman, *Towards an Integrated Theory of Intellectual Property*, 88 VA. L. REV. 1455, 1494 (2002).

[32] *See infra* Section II.B.

[33] *See* Andrew Beckerman-Rodau, *The Choice Between Patent Protection and Trade Secret Protection: A Legal and Business Decision*, 84 J. PAT. & TRADEMARK OFF. SOC'Y 371, 395–96 (2002); Roy E. Hofer & L. Ann Fitzgerald, *New Rules for Old Problems: Defining the Contours of the Best Mode Requirement in Patent Law*, 44 AM. U. L. REV. 2309, 2337 (1995).

[34] *See infra* Part II.

[35] *See infra* Part II. By "distinct product markets," we generally refer to the usage in the Supreme Court's opinion in *Jefferson Parish Hospital District No. 2 v. Hyde*, which found separate markets to exist if there is enough consumer demand for the products to be sold separately. 466 U.S. 2, 21–22 (1984), *overruled on other grounds by* Ill. Tool Works Inc. v. Indep. Ink, Inc., 547 U.S. 28 (2006).

[36] *See infra* Parts I–II (discussing the standard explanation of patents and trade secrets as substitutes and introducing and describing "data-generating patents" respectively).

[37] *See infra* Part II. Other scholars have recognized that patents on medical diagnostic tests may allow for the control of information generated by the patented tests as a trade secret. *See* Dan L. Burk, *Patents as Data Aggregators in Personalized Medicine,* 21 B.U. J. SCI. & TECH. 233, 233–49 (2015) (providing a general theory of how patent holders in the field of personalized medicine can generate massive databases of patient information protected by trade secret law); John M. Conley et al., *Myriad*

under which data-generating patents may be socially detrimental, and offer and analyze a variety of potential solutions.[38]

I. PATENTS AND TRADE SECRETS: FROM SUBSTITUTES TO COMPLEMENTS

Patents and trade secrets are traditionally considered economic substitutes: if one decides to patent an invention, one must disclose the details of the invention and forgo trade secret protection.[39] However, a handful of scholars have recognized that patents and trade secrets may sometimes act as economic complements, because patent law does not always require full disclosure of the invention.[40] In order to place our description of data-generating patents in a suitable theoretical context, we briefly describe these two views of the relationship between patents and trade secrets.

A. The Majority View: Patents as "Substitutes" for Trade Secrets

The standard view portrays patents and trade secrets as economic substitutes, namely two mutually exclusive forms of intellectual property protection from which an innovator must ultimately select only one for a given invention.[41] In short, commentators have largely suggested that

After Myriad*: The Proprietary Data Dilemma*, 15 N.C. J.L. & TECH. 597, 616–17 (2014) (describing Myriad's proprietary database strategy); Robert Cook-Deegan et al., *The Next Controversy in Genetic Testing: Clinical Data as Trade Secrets?*, 21 EUR. J. HUM. GENETICS 585 (2013) (cataloguing concerns regarding the retention of information generated by medical diagnostic tests as trade secrets). However, unlike the treatment in this Article, these scholars have limited their inquiry to the area of personalized medicine. *See* Burk, *supra*, at 233–49; Cook-Deegan et al., *supra*, at 585–87. In this regard, while Burk's analysis of Myriad's patented genetic tests is similar in some ways to ours, we conceived of, publicly presented, and drafted that portion of this Article prior to the presentation or dissemination of Burk's article. *See* Dennis Crouch, *Twenty Thoughts on the Importance of* Myriad, PATENTLY-O (June 14, 2013), http://patentlyo.com/patent/2013/06/myriad.html [https://perma.cc/WR4K-UECF] ("Brenda Simon . . . notes that *Myriad* 'does not impact one of the most valuable aspects made possible through Myriad's patent protection: a private biobank of patient data containing information about additional mutations that Myriad can maintain as a trade secret.'"); Ted Sichelman & Brenda Simon, Generating Trade Secrets From Patents, Presentation at the 14th IP Scholars Conference, Berkeley, California (Aug. 7, 2014). More importantly, in contrast to these biomedical-focused works, we introduce the broad concept of "data-generating patents" and explain how they apply across numerous technology types and industries. *See infra* Sections II.A–C. Additionally, unlike the prior literature, we provide a general analytical framework to distinguish problematic from unproblematic data-generating patents and offer and assess numerous prescriptive options to remedy any potentially detrimental effects. *See infra* Part III.

38 *See infra* Part III.

39 *See infra* Section I.A.

40 *See infra* Section I.B.

41 *See* J. Jonas Anderson, *Secret Inventions*, 26 BERKELEY TECH L.J. 917, 923–24 (2011); Mark A. Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, 61 STAN. L. REV. 311, 314

patenting essentially destroys trade secret protection, because patents nominally require disclosure of the underlying invention sufficient to enable one of skill in the art to make and use the invention.[42] Gideon Parchomovsky and Peter Siegelman aptly describe the traditional view: "As a substitute for patent protection, trade secrecy presents businesses with a choice between patent and trade secret protection. While firms can elect either option, they cannot employ both modes to protect the same information."[43]

More precisely, patents are an intellectual property right that prevents others from making, using, importing, selling, or offering for sale the patented invention for a period of twenty years from the date of patent application.[44] In exchange for this exclusionary right, the so-called "quid pro quo" of patent law is the disclosure of the invention in the patent document so as to increase the storehouse of public knowledge.[45] The typical rationale for such an exchange is utilitarian: in order to promote innovation, society is willing to impart a limited-time legal monopoly to the inventor in exchange for, at least in part, the knowledge about the invention, freely available to the public after the term of the patent has expired.[46]

The patent–trade secret substitute story is nuanced, however. The types of subject matter that may be patented are both broader and narrower than what can be covered by a trade secret.[47] In essence, any commercially valuable information that may be kept secret can be eligible for trade secrecy protection.[48] Patentable subject matter, on the other hand, consists

(2008); Andrew A. Schwartz, *The Corporate Preference for Trade Secret*, 74 OHIO ST. L.J. 623, 624 (2013).

42 *See* 35 U.S.C. § 112 (2012); *see also* Parchomovsky & Siegelman, *supra* note 31, at 1494. To be certain, nobody disputes that trade secrecy may be used to protect the ideas and information *prior to* the patenting process. *See* Sichelman & Graham, *supra* note 29, at 136 n.137. The question of primary theoretical concern, however, is whether trade secrecy continues to play a role after patenting. *See id.*

43 Parchomovsky & Siegelman, *supra* note 31, at 1494.

44 §§ 271, 154(a)(2).

45 Universal Oil Prods. Co. v. Globe Oil & Ref. Co., 322 U.S. 471, 484 (1944); *see also* Lisa Larrimore Ouellette, *Do Patents Disclose Useful Information?*, 25 HARV. J.L. & TECH. 545, 546 (2012).

46 *See* Peter S. Menell, *Intellectual Property: General Theories*, *in* 2 ENCYCLOPEDIA OF LAW AND ECONOMICS 129, 130–33 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000).

47 *See* Oren Bar-Gill & Gideon Parchomovsky, *Law and the Boundaries of Technology-Intensive Firms*, 157 U. PA. L. REV. 1649, 1685 (2009).

48 The Uniform Trade Secrets Act (UTSA) defines a "trade secret" as:

> [I]nformation, including a formula, pattern, compilation, program, device, method, technique or process, that: (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.

only of a "process, machine, manufacture, or composition of matter" that does not constitute an abstract idea, law of nature, or natural phenomena, regardless of whether the invention can be kept secret if used commercially.[49] Moreover, the claimed invention must also be novel and nonobvious,[50] whereas old and obvious subject matter may be covered by a trade secret as long as it is not publicly known.[51]

So patents and trade secrets are only substitutes to the extent that the subject matter can either be patented or potentially kept a secret.[52] For this overlapping subject matter, the substitute theory posits that the description of the invention in the patent eliminates any protection under trade secret law.[53] Specifically, a patent application must disclose sufficient detail (1) to satisfy the written description requirement that the inventor demonstrate "possession" of the claimed invention and (2) to enable a person of skill in the art to make and use the invention.[54] As a consequence of these disclosure requirements—coupled with trade secret law's requirement to maintain any protected information as secret—under the substitution theory, one may not obtain both trade secret and patent rights.[55]

UNIF. TRADE SECRETS ACT § 1(4) (UNIF. LAW COMM'N, amended 1985); *see also* ROGER M. MILGRIM & ERIC E. BENSEN, MILGRIM ON TRADE SECRETS § 1.01 (Matthew Bender & Co. ed. 2015).

[49] § 101; *see also* Mayo Collaborative Servs. v. Prometheus Labs., Inc., 566 U.S. 66, 72 (2012) (excluding inventions from eligibility unless they are sufficiently applied).

[50] §§ 102, 103; DONALD S. CHISUM, CHISUM ON PATENTS §§ 3.01, 5.01 (Matthew Bender & Co. 2013).

[51] *See, e.g.*, JAY DRATLER, JR., INTELLECTUAL PROPERTY LAW: COMMERCIAL, CREATIVE, AND INDUSTRIAL PROPERTY § 4.03 (Law Journal Seminars-Press 2015); Jason Mazzone & Matthew Moore, *The Secret Life of Patents*, 48 WASHBURN L.J. 33, 41–42 (2008).

[52] *Cf.* Katherine J. Strandburg, *What Does the Public Get? Experimental Use and the Patent Bargain*, 2004 WIS. L. REV. 81, 103–04 ("[I]ncreasing the effectiveness of disclosure will be unlikely to have a significant impact on incentives to invent because disclosure has an inherently greater impact on inventions that could have been maintained as trade secrets.").

[53] *See* Sichelman & Graham, *supra* note 29, at 136.

[54] § 112(a).

[55] Moreover, there is a legal incentive to obtain patent protection rather than maintain trade secrets, because (at least historically) a prior inventor using the invention as a trade secret would not invalidate, but would infringe, the patent of a later inventor on the same invention. *See* W.L. Gore & Assocs., Inc. v. Garlock, Inc., 721 F.2d 1540, 1550 (Fed. Cir. 1983); Metallizing Eng'g Co. v. Kenyon Bearing & Auto Parts Co., 153 F.2d 516, 520 (2d Cir. 1946). Following passage of the America Invents Act (AIA), some commentators have argued that the *Metallizing* doctrine has been eliminated, though the majority scholarly view appears to be that it persists. *See* Mark A. Lemley, *Does "Public Use" Mean the Same Thing It Did Last Year?*, 93 TEX. L. REV. 1119, 1122–24 (2015); Dmitry Karshtedt, *The Riddle of Secret Public Use: A Response to Professor Lemley*, 93 TEX. L. REV. SEE ALSO 159, 161 (2015). Everyone agrees the AIA has erected a limited prior user defense to infringement for trade secrets used commercially in a manufacturing process more than one year prior to the filing of an applicable patent filed after September 16, 2011, but such a narrow defense is unlikely to change the patent–trade secret decision calculus in most instances. § 273 (expanding the prior use defense under previous law).

The substitution model has many implications for intellectual property theory and practice. For instance, which form of protection the inventor seeks out depends upon the salient features and consequences of patent and trade secret protection to the invention at hand.[56] Inventors and firms will typically consider the duration of protection, likelihood of reverse engineering, likelihood of independent invention, detectability of infringement, and the cost of procuring and enforcing the rights inherent in the protection sought.[57] For example, patent protection may be an easier choice for mechanical and electrical inventions because they have generally been more prone to reverse engineering than other types of inventions, such as manufacturing processes for biologics.[58] These and other details of substitution theory, however, are not particularly important for our discussion here. As such, we turn next to critiques of the theory.

B. The Minority View: Patents and Trade Secrets as Complements

Several scholars, including one of us, have previously questioned the standard view that patents and trade secrets are merely substitutes.[59] Despite the black letter rule that an inventor "can lawfully claim only what he has invented and described,"[60] the U.S. Patent & Trademark Office (USPTO) and the courts typically allow patent claims that are much broader than what is actually disclosed in a patent application.[61]

---

[56] *See* David D. Friedman, William M. Landes & Richard A. Posner, *Some Economics of Trade Secret Law*, 5 J. ECON. PERSP. 61, 62–63 (1991); Note, *Patent Preemption of Trade Secret Protection of Inventions Meeting Judicial Standards of Patentability*, 87 HARV. L. REV. 807, 821 (1974).

[57] *See* Beckerman-Rodau, *supra* note 33, at 377–401; Kimberly A. Moore, *Worthless Patents*, 20 BERKELEY TECH. L.J. 1521, 1545 (2005); Schwartz, *supra* note 41, at 630–31.

[58] *See, e.g.*, ANDREW RUDGE, GUIDE TO EUROPEAN PATENTS § 8:4 (2008) ("Just as chemical products reveal their secrets on chemical analysis, so mechanical products can reveal their mechanism on dismantling or reverse-engineering, and the sequence of commands behind a computer programs [sic] can be decoded."); Robert P. Merges, *Rent Control in the Patent District: Observations on the Grady-Alexander Thesis*, 78 VA. L. REV. 359, 376 (1992) ("It is well-known, for instance, that process inventions are easier to keep secret than product inventions."); Daniel C. Munson, *The Patent-Trade Secret Decision: An Industrial Perspective*, 78 J. PAT. & TRADEMARK OFF. SOC'Y 689, 696–98 (1996) (noting that electrical and mechanical inventions are often easy to reverse engineer); W. Nicholson Price II, *Regulating Secrecy*, 91 WASH. L. REV. (forthcoming 2017) (discussing trade secret protection for manufacturing processes of patented biologics).

[59] *See* Sichelman & Graham, *supra* note 29, at 136.

[60] O'Reilly v. Morse, 56 U.S. (1 How.) 62, 121 (1853); *see* CHISUM, *supra* note 50, § 1.03[2][b]; *see also* Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265, 268 (1977).

[61] *See* JAMES BESSEN & MICHAEL J. MEURER, PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK 66–68 (2008) (describing biotechnology and software patent claims that greatly exceeded the scope of the disclosed invention); Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495, 1499–1500 (2001) (explaining how the USPTO is unable to adequately examine each patent and grants many broad patents as a result).

Specifically, a patent will usually disclose just one or a few tangible "embodiments" of the invention in the patent's specification, but will often claim thousands of different embodiments in a claim.[62] This result is due in substantial part to the courts' and Patent Office's relatively lax implementation of the written description and enablement requirements.[63] For instance, Myriad's patents related to the BRCA1 and BRCA2 genes covered numerous cancer-associated variants of these genes but disclosed only a small subset of those variants in the patent specifications themselves.[64]

More recently, disclosure standards have been watered down even further by the essential elimination of the "best mode" requirement from patenting, which requires the inventor to disclose those details—such as temperature and pressure ranges, height and width, and similar parameters—that optimize the invention.[65] However, under the recently passed America Invents Act, although best mode is still a nominal requirement for patenting, failure to disclose it can no longer serve as a basis to invalidate or render a patent unenforceable.[66] Because a patent

---

[62] Indeed, Jeffrey Lefstin views modern claims as potentially covering "an infinite variety of embodiments." Jeffrey A. Lefstin, *The Formal Structure of Patent Law and the Limits of Enablement*, 23 BERKELEY TECH. L.J. 1141, 1169 (2008).

[63] *See* Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155, 1168 (2002); Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 845–46 (1990).

[64] U.S. Patent Nos. 5,747,282 (issued May 5, 1998); 5,693,473 (issued Dec. 2, 1997); 5,837,492 (issued Nov. 17, 1998). To be certain, disclosure standards are often more rigorous in the biomedical fields, especially in the courts. *See* Ariad Pharm., Inc. v. Eli Lilly & Co., 598 F.3d 1336 (Fed. Cir. 2010). Nonetheless, as the Myriad patents illustrate, even in these fields, a patent disclosure typically contains far fewer examples than what is covered by the claims. *See, e.g.*, Sean B. Seymore, *The Teaching Function of Patents*, 85 NOTRE DAME L. REV. 621, 637 (2010) (describing problems with disclosure, including how patent applicants have "develop[ed] various claim drafting schemes so as to maximize the breadth of a claim based on certain illustrative, or sometimes a modicum of, disclosure" in the specification (quoting C. Leon Kim, *Transition from Central to Peripheral Definition Patent Claim Interpretation System in Korea*, 77 J. PAT. & TRADEMARK OFF. SOC'Y 401, 404 (1995))).

[65] 35 U.S.C. § 112 (2012) (The inventor is to "set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention."); *see* Donald S. Chisum, *Best Mode Concealment and Inequitable Conduct in Patent Procurement: A Nutshell, a Review of Recent Federal Circuit Cases and a Plea for Modest Reform*, 13 SANTA CLARA COMPUTER & HIGH TECH. L.J. 277, 278 (1997) (stating that the best mode requirement requires inventors "reveal to the public the preferred implementations of their inventions and . . . information that is material to the patentability of an invention"); *see also* Applied Materials, Inc. v. Advanced Semiconductor Materials Am., Inc., 98 F.3d 1563, 1581 (Fed. Cir. 1996) (discussing how inventors can "disclose only minimal parts of their inventions" and "thereby hide the commercial value that resides in the best mode of practicing their inventions and gain the benefit of both the exclusionary right of the patent and the 'quasi trade secret' of the best mode").

[66] 1 LESTER HORWITZ & ETHAN HORWITZ, PATENT OFFICE RULES AND PRACTICE § 71[H] (Matthew Bender & Co. 2015) ("Section 15 of the Leahy-Smith America Invents Act does not eliminate the [best mode] requirement in 35 U.S.C. § 112 . . . but does amend 35 U.S.C. § 282 . . . to provide that the failure to disclose the best mode shall not be a basis on which any claim of a patent

examiner will generally be unable to detect failure to disclose best mode, it is on some accounts largely a dead letter and, at best, severely weakened in effect.[67]

The result of these fairly weak disclosure rules is that a patent applicant can often secure a patent without disclosing all of the technologically and commercially important details of an invention. This withheld knowledge can in turn be protected under trade secret law.[68] One very common example of this practice is that inventors can typically retain software source code as a trade secret, yet still obtain a patent.[69] In another example, Pilkington Glass used patents and trade secrets to protect codified and tacit elements of the company's "float glass" invention, a radical improvement in creating smooth glass.[70] Thus, contrary to the standard view, patents and trade secrets can simultaneously serve as economic complements.[71]

Additionally, an inventor need not—in fact, cannot—update the disclosure during the prosecution of the patent application in front of the Patent Office.[72] Thus, if an inventor files a patent early in the R&D process, the disclosed best mode of making and using the invention may greatly differ from the ultimate commercial embodiment.[73] So even if inventors

may be canceled or held invalid or otherwise unenforceable."). *See generally* Brian J. Love & Christopher B. Seaman, *Best Mode Trade Secrets*, 15 YALE J.L. & TECH. 1, 3 (2012) (explaining that the AIA may yield the unintended securing of protection through both patent and trade secret law).

67 Marc A. McClain, *Who Are the New "Best Mode" Police? An Analysis of Proposed New Methods of Enforcement of the Best Mode Requirement After the Leahy-Smith America Invents Act*, 46 U. TOL. L. REV. 191, 219 (2014).

68 *See* Stuart J.H. Graham & Ted Sichelman, *Why Do Start-Ups Patent?*, 23 BERKELEY TECH. L.J. 1063, 1086–88 (2008) ("[B]ecause the enablement and written description requirements are weak, . . . a patentee may often be able to patent an invention and keep its 'secret sauce' a trade secret.").

69 *See* Gregory J. Maier, *Software Protection—Integrating Patent, Copyright and Trade Secret Law*, 69 J. PAT. & TRADEMARK OFF. SOC'Y 151, 163–65 (1987) (noting that patent law does not require the disclosure of software code).

70 *See* United States v. Pilkington PLC, No. CV 94-345, 1994 WL 750645 (D. Ariz. Dec. 22, 1994).

71 *See* Douglas Lichtman, *How the Law Responds to Self-Help*, 1 J.L. ECON. & POL'Y 215, 230 (2005) ("Trade secret law thus fills an important void, creating that incentive and thereby acting as both a complement to and competitor for patent law."); Michael J. Meurer & Craig Allen Nard, *Invention, Refinement and Patent Claim Scope: A New Perspective on the Doctrine of Equivalents*, 93 GEO. L.J. 1947, 1974 (2005) ("First, prosecutors and inventors strategically choose narrow claims . . . because the inventor decides to seek a mix of patent and trade secret protection."); Elisabetta Ottoz & Franco Cugno, *Choosing the Scope of Trade Secret Law When Secrets Complement Patents*, 31 INT'L REV. L. & ECON. 219 (2011) (presenting an economic model wherein trade secrets and patents can serve as complements).

72 *See* 35 U.S.C. § 132 (2012).

73 *See* Robert P. Merges, *Commercial Success and Patent Standards: Economic Perspectives on Innovation*, 76 CALIF. L. REV. 803, 807 (1988) ("[T]he innovation will in all likelihood be different in significant respects from the invention due to the changes necessary to turn the invention into a

disclose all of the details of the invention at the time of filing, they may retain later discovered details—such as optimal working conditions—as a trade secret.[74] In fact, knowing as much—and given the weak disclosure requirements in the first instance—a common strategy is to file early on the concept in order to retain commercial know-how learned after filing as a trade secret.[75] Indeed, inventions need not be commercialized to garner protection under the patent laws.[76] Instead, an adequate written disclosure of the invention suffices to "reduc[e] the invention to practice."[77] In this regard, there is often a wide gulf between a patent disclosure that can satisfy the written description, enablement, and best mode disclosure requirements[78] and the documents that are actually used to build commercial products.[79] Thus, as a patented product or method evolves through the commercialization cycle, information generated by its owner, sellers, and consumers about the invention can be retained as a trade secret, again supporting the complements view.[80]

II. THE USES AND UNIQUENESS OF DATA-GENERATING PATENTS

The use of patents and trade secrets as complements discussed in Part I focused on information about the *invention itself* that could be retained as a secret.[81] For instance, a patentee may keep secret the fact that its patented process for baking clay works best at 451 degrees Fahrenheit or that a

commercial product."); Dennis Crouch, *The Trade Secret Value of Early Patent Filing*, PATENTLY-O (Oct. 23, 2008), http://www.patentlyo.com/patent/2008/10/the-trade-secre.html [https://perma.cc/HF7D-49JF] ("[M]any if not most patent applications are filed well before the associated product or method is ready for public consumption – before the inventor knows the best *commercially viable* mode.").

[74] *See* Chisum, *supra* note 65, at 283; Steven B. Walmsley, *Best Mode: A Plea to Repair or Sacrifice This Broken Requirement of United States Patent Law*, 9 MICH. TELECOMM. & TECH. L. REV. 125, 134 (2002).

[75] *Cf.* Transco Prods. Inc. v. Performance Contracting, Inc., 38 F.3d 551, 557–58 (Fed. Cir. 1994) (Rich, J.) (noting the benefits of filing early and avoiding disclosure of improvements of the invention).

[76] Cont'l Paper Bag Co. v. E. Paper Bag Co., 210 U.S. 405, 429 (1908) (finding that a patent is not unenforceable merely because the patentee neglected to put it to commercial use and that it is "the privilege of any owner of property to use or not use it, without question of motive"); Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341, 355 (2010) ("[T]he patent laws do not require inventions to be in a commercialized form to garner protection.").

[77] *See* Christopher A. Cotropia, *The Folly of Early Filing in Patent Law*, 61 HASTINGS L.J. 65, 128 (2009).

[78] 35 U.S.C. § 112 (2012).

[79] *See* CFMT, Inc. v. YieldUp Int'l Corp., 349 F.3d 1333, 1338 (Fed. Cir. 2003) ("Enablement does not require an inventor to meet lofty standards for success in the commercial marketplace.").

[80] *See* Parchomovsky & Siegelman, *supra* note 31, at 1494 ("[T]rade secret law applies to unpatentable as well as patentable information. Consequently, trade secrecy serves both as a complement to and substitute for patent protection.").

[81] *See supra* Section I.B.

customer determined that the clay is optimally heated for four minutes and thirty-three seconds.[82] Surely inventors have gathered information about patented inventions and subsequently kept it secret since the origins of the patent system.[83]

"Data-generating patents" differ from the traditional use of patents and trade secrets as complements in that the inventions covered by these patents—by definition—often generate information apart from the invention itself. In economic terms, data-generating inventions may yield information in distinct product markets.[84] This information—which is typically protectable as a trade secret—may be used to improve the invention itself or in an entirely different manner.[85] For instance, search-engine technology may generate data about its users, which can then be used to improve the operation of the search engine or to determine market trends, or for any number of other uses, such as advertising to users.[86]

Data-generating inventions per se are not new. Indeed, the invention of various writing instruments and research tools throughout history has generated unfathomable amounts of information unrelated to the invention itself.[87] Nonetheless, our notion of data-generating patents is a newly recognized form of patents and trade secrets functioning as economic complements.[88] In this regard, data-generating patents have become of particular note in the current era of big data, in which technology now allows for low-cost data storage, retrieval, and analysis.[89] Whereas the

---

[82] N. Telecom Ltd. v. Samsung Elecs. Co., 215 F.3d 1281, 1297 (Fed. Cir. 2000) (holding that an inventor need not disclose information that is not related to the operation of the claimed invention to satisfy the best mode requirement).

[83] *See* Parchomovsky & Siegelman, *supra* note 31, at 1494; Ted Sichelman et al., *Retracing the Origins of the Patent System* (unpublished manuscript) (on file with authors) (stating that public records of petitions for patents in the Venetian system contained concise summaries of the invention that were not generally updated during examination).

[84] *See* Thomas M. Jorde & David J. Teece, *Rule of Reason Analysis of Horizontal Arrangements: Agreements Designed to Advance Innovation and Commercialize Technology*, 61 ANTITRUST L.J. 579, 614 (1993) ("If there is no substitution to the new product, then the products represent distinct markets.").

[85] *See infra* Section II.A.

[86] *See infra* Section II.A.

[87] *See, e.g.*, U.S. Patent Nos. 3,013,467 (filed Nov. 7, 1957) (issued Dec. 19, 1961) (disclosing electronic microscope); 6,934,574 (filed June 20, 2002) (issued Aug. 23, 2005) (disclosing MRI scanner).

[88] *See supra* Section I.B.

[89] *See, e.g.*, Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 536 (2014) ("[M]any experts believe that [big data] will soon lead the way to new frontiers in science and innovation." (internal citations omitted)); W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J. LAW & TECH. 419, 436 (2015) ("The wealth of data available in electronic health records of patients suffering from different ailments and responding to drugs they take for other purposes may be mined by big-data algorithms, which can suggest new uses."); Nicolas P. Terry, *Big Data Proxies and Health Privacy*

Library of Alexandria could at best catalog its scrolls by subject, author, title, or a few lines of text—a method of indexing prevalent until fairly recently[90]—Google and other search engines can effectively catalog not only at the granular level of words and phrases, but also by the very concepts represented by those words and phrases.[91] Whereas Ben Franklin's yearly updated almanac catalogued and analyzed hundreds of pages of information for practical use,[92] artificially intelligent algorithms can now scan and analyze millions of pages in seconds and draw new associations from them.[93]

These rapid changes in the nature and uses of data present unique social and economic concerns, particularly for innovation and competition.[94] These concerns are amplified when patent protection is used to provide legal exclusivity—which can often lead to market power—for a data-generating invention.[95] In particular, the owner of a market power-conferring, data-generating patent can use its legal rights to enjoy an advantage not only in the product market of the invention, but also in the market for the information generated by the invention. In this regard, unlike trade secrecy, reverse engineering or independent invention is no defense to patent infringement. Thus, data-generating patentees are presented with no

---

Exceptionalism, 24 HEALTH MATRIX 65, 77 (2014) ("Big data refers not only to the collection and storage of extremely large data sets but also the data mining and predictive analytic routines that process the data . . . .").

[90] *See* ANN BLAIR, TOO MUCH TO KNOW: MANAGING SCHOLARLY INFORMATION BEFORE THE MODERN AGE 16 (2010); WILLIAM S. PIERCE, FURNISHING THE LIBRARY INTERIOR 68 (1980).

[91] *See generally* Cynthia L. Barrett, Book Review, 2 NAT'L ACAD. OF ELDER LAW ATTORNEYS J. 7 (2006) (reviewing ANDREW WEIL, HEALTHY AGING (2005)); JOSEPH B. MILLER, INTERNET TECHNOLOGIES AND INFORMATION SERVICES 372 (2d ed. 2014).

[92] *See, e.g.*, RICHARD SAUNDERS, POOR RICHARD, 1733: AN ALMANACK (Benjamin Franklin prtg. 1732) (spanning less than thirty pages).

[93] *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 84 (2013); Stephen F. DeAngelis, *Artificial Intelligence: How Algorithms Make Systems Smart*, WIRED (Sept. 5, 2014), http://www.wired.com/insights/2014/09/artificial-intelligence-algorithms-2/ [https://perma.cc/CUG7-KVWV].

[94] *See infra* Part III.

[95] *See infra* Part III. By "market power," we refer to the standard understanding that such power is measured by a firm's ability to increase a product's prices "without losing a significant fraction of its sales." Benjamin Klein, *Market Power in Antitrust: Economic Analysis After* Kodak, 3 SUP. CT. ECON. REV. 43, 44 (1993); *see also* Thomas G. Krattenmaker & Steven C. Salop, *Anticompetitive Exclusion: Raising Rivals' Costs to Achieve Power over Price*, 96 YALE L.J. 209, 212–13 (1986). Although there is some disagreement over exactly what this understanding entails, our analysis here would not change under alternative definitions. *Cf.* Stephen Yelderman, *Do Patent Challenges Increase Competition?*, 83 U. CHI. L. REV. (forthcoming 2017) (manuscript at 17) (on file with authors) (questioning "the proposition that patents always or even typically confer market power"). *See generally* Alan J. Daskin & Lawrence Wu, *Observations on the Multiple Dimensions of Market Power*, ANTITRUST, Summer 2005, at 53.

risk of independent discovery or reverse engineering while the patent on the underlying invention is in effect. Moreover, once the patent term ends, or even if the patent is invalidated, the holder continues to enjoy legal exclusivities afforded by trade secret law and its potentially indefinite term of protection. Beyond its legal rights, the data-generating patentee may also benefit from the de jure lead time advantage secured by the prior patent in its compilation of data. These potential economic consequences of data-generating patents were not contemplated in the delicate balance erected by the patent system and, as we explain below, can—like big data more generally—potentially cause major social and economic harm.

Before turning to these potential harms, we describe the unique nature of data-generating patents in more detail. Data-generating patents appear in areas ranging from genetic testing and medical devices to internet search and social networking. In some circumstances, the data acquired, aggregated, and analyzed through the use of the patented invention may be more valuable than the patent itself. The first Section of this Part describes the use of data-generating patents in two areas: genetic testing and internet search. It then briefly describes many other technological areas in which data-generating patents play a role. We conclude by explaining how data-generating patents function as a form of intellectual property in which patents and trade secrets function as economic complements in a unique manner. In contrast to previously recognized examples, data-generating patents may yield trade secrets in different product markets and well after the term of the patent expires or the patent is invalidated.

A. The Rise of Data-Generating Patents

1. From Patents to Proprietary Databases of Genetic Disease Markers.—The value of trade secrets obtained through data-generating patents is particularly evident in the area of genetic testing, particularly the generation of proprietary databases of patients' genetic information derived from patented diagnostic tests. Most notably, Myriad Genetics had been the sole provider in the United States of testing for the BRCA1 and BRCA2 genes, which are markers for breast cancer.[96] Myriad collects and stores information from its patients about variations,

[96] *See infra* note 99; *see also* Ass'n for Molecular Pathology v. U.S. Patent & Trademark Office, 702 F. Supp. 2d 181, 189 (S.D.N.Y. 2010), *aff'd in part, rev'd in part*, 689 F.3d 1303 (Fed. Cir. 2012), *aff'd in part, rev'd in part sub nom*, Ass'n for Molecular Pathology v. Myriad Genetics, Inc., 133 S. Ct. 2107 (2013); Ashley McHugh, Note, *Invalidating Gene Patents:* Association for Molecular Pathology v. U.S. Patent & Trademark Office, 62 HASTINGS L.J. 185, 187 (2010).

phenotypes, populations, and family histories.[97] Over 1.5 million patients have used and contributed data to Myriad's BRCA testing services.[98]

Until recently, Myriad's status as the sole testing provider for these genes was based, at least in part, on its strong assertion of its patent rights.[99] Before the Supreme Court's recent decision invalidating some of its patent claims, Myriad had the exclusive ability to collect clinical data from patients as a result of its patent protection for over a decade, maintaining their information in a private database.[100]

Myriad usually provides three types of test results when comparing a patient's genetic information with a naturally occurring genetic sequence: no variations, harmless variations, or mutations that are clearly harmful.[101] For the first two types, no variations or harmless variations, the difference between the patient's sequence and the naturally occurring sequence does not indicate an increased susceptibility to breast or ovarian cancer.[102] For the third type, where the mutations are clearly harmful, the patient is at increased risk for developing cancer.[103]

Occasionally, however, the test results are more complicated to understand.[104] For variants of unknown significance (VUS), the effects of the variations are difficult to interpret, so it is unclear whether the patient is at increased risk of cancer.[105] Myriad states that the portion of patients with

---

97 *See* Cook-Deegan et al., *supra* note 37, at 585 (describing Myriad's "extensive database that relates variants of uncertain significance to phenotype, details their frequency in various populations and includes genetic studies on patient families").

98 *See* *History*, MYRIAD, https://www.myriad.com/about-myriad/inside-myriad/history/ [https://perma.cc/4DZ6-TGM2].

99 *See* Cook-Deegan et al., *supra* note 37, at 585 ("[Myriad's] status as the sole commercial provider of BRCA testing in the United States is a consequence of its exclusive US patent rights."). Myriad's vigorous enforcement strategy was unusual among patent holders of biomedical technologies at the time. *See* Christopher Heaney et al., *The Perils of Taking Property Too Far*, 1 STAN. J.L. SCI. & POL'Y 46 (2009); Christopher M. Holman, *The Impact of Human Gene Patents on Innovation and Access: A Survey of Human Gene Patent Litigation*, 76 UMKC L. REV. 295, 299 (2007); Brenda M. Simon, *Patent Cover-Up*, 47 HOUS. L. REV. 1299, 1308 (2011).

100 *Myriad Genetics*, 133 S. Ct. at 2111–14.

101 *See* Cook-Deegan et al., *supra* note 37, at 585.

102 *See generally* CTR. FOR GENETICS EDUC., UNDERSTANDING GENETIC TESTS FOR BREAST AND OVARIAN CANCER THAT RUNS IN THE FAMILY: INFORMATION AND DECISION AID 20 (2014), http://www.genetics.edu.au/Publications-and-Resources/PublicationsBrochuresandPamphlets/Understanding%20genetic%20tests%20for%20Breast%20and%20Ovarian%20Cancer%20that%20runs%20in%20the%20family-Information%20and%20Decision%20Aid.pdf [https://perma.cc/V26N-M7JS].

103 *See id.* at 21.

104 *See* Cook-Deegan et al., *supra* note 37, at 585 (explaining that "variants of unknown significance" are "difficult to interpret"). *See generally* CTR. FOR GENETICS EDUC., *supra* note 102, at 21 (explaining that if a patient receives an uncertain variant result, "it is not clear currently if the variation is harmless, or if it is a variation that is making the gene faulty").

105 Cook-Deegan et al., *supra* note 37, at 585.

unknown variations is only 3%.[106] For most BRCA testing carried out by other companies in Europe, where Myriad's patents have not been strongly enforced, the portion of cases resulting in unknown variations is as high as 20%.[107] Myriad's lower unknown variation rate results from its ability to interpret its results in light of its very large, proprietary database.[108] Myriad builds and maintains the database at its own expense, and offers free testing to some families of consumers with unknown variations to try to determine the significance of the variants.[109] As Myriad acquires additional information, it will have fewer variants of unknown significance.[110] In contrast, competitors using public databases would not have access to Myriad's proprietary information in interpreting patients' test results.[111]

Thus, Myriad's patents arguably have provided it a competitive advantage in generating a database of mutations and other clinical information that will be difficult and costly for competitors to replicate.[112] Even if all competitors cooperated, contributing their data to a public or quasi-public database would not appear to be a viable alternative.[113] In other words, so-called "private ordering" solutions to the costs of data-generating patents may not always be feasible.[114] Such barriers, coupled with the inability for the market to self-correct and lack of regulatory intervention,

---

106 *See id.*

107 *Id.*

108 *See id.* at 586 (noting that to help its patients, Myriad uses "its database to reduce the frequency with which it reports a VUS"); *id.* ("[O]utsiders do not have access to Myriad's database."); Jacob S. Sherkow & Christopher Scott, *Myriad Stands Alone*, 32 NATURE BIOTECHNOLOGY 620 (2014) (describing Myriad's proprietary database).

109 *See* Cook-Deegan et al., *supra* note 37, at 586.

110 *See id.*

111 *See id.*

112 *See id.* (stating that recreating the database would be "redundant and thus expensive").

113 *See* Conley et al., *supra* note 37, at 597, 614–15 ("Myriad has used its patent-based monopoly as the sole BRCA 1 and 2 test provider to develop, at its own cost, an extensive database that relates VUSs to phenotypes, details the frequency of VUSs in various populations, and includes genetic studies on patient families. There is no comparable public database."); *cf.* Angela M. Oliver, *Personalized Medicine in the Information Age: Myriad's De Facto Monopoly on Breast Cancer Research*, 68 SMU L. REV. 537, 551–52 (2015) ("If researchers could create a comparable database through such efforts, it would strip Myriad of trade secret protection for its database.").

114 *See, e.g.*, Jay P. Kesan, *Economic Rationales for the Patent System in Current Context*, 22 GEO. MASON L. REV. 897, 918–19 (2015) ("Private ordering . . . refers to 'circumstances where parties, given extant legal and regulatory regimes, order the substance of their affairs and transactions as they see fit and resort to the judicial system for enforcement.'" (quoting F. Scott Kieff & Troy A. Paredes, *Engineering a Deal: Toward a Private Ordering Solution to the Anticommons Problem*, 48 B.C. L. REV. 111, 114 n.15 (2007))).

afford Myriad the ability to extend its market power well beyond the expiration or invalidation of its patents.[115]

2. *Internet Search as a Personalized, Proprietary Process*.—Another technological field in which data-generating patents are valuable is internet search, in which Google is currently the market leader.[116] Like Myriad's patents, the power to exclude competitors helped enable Google to amass a huge set of proprietary data about its users, which is in turn protected by trade secret law.[117] Google has leveraged this data not only to improve its underlying search technology but also to excel in secondary markets, critically including targeted advertising, from which it earns the bulk of its revenue.[118]

Commentators have claimed, however, that Google's patents are essentially irrelevant to its market dominance. For instance, Dan Burk contends that "even if the search algorithm . . . is patented, data aggregates around search engine operation by virtue of the search function, . . . not as a result of legal exclusivity."[119] There are two major reasons to dispute such claims. First, one must distinguish between the *mechanism for aggregating data*—e.g., a "search function" or "diagnostic screen"—and the *economic power* a data-generating patent holder may enjoy to draw more users to its data-aggregating mechanism and, hence, generate large amounts of data. In this regard, although there are many reasons a firm may enjoy market power, as long as patents play a substantial role in such power, they present a concern worth discussing.

[115] *See* Burk, *supra* note 37, at 253–54 (stating that if competitors are unable "to enter the market at the end of the patent period, monopoly pricing might be maintained during the patent period and beyond").

[116] George Slefo, *Microsoft and Yahoo Search Share Grows but Still Trails Google by Miles*, ADVERTISINGAGE (Dec. 23, 2015), http://adage.com/article/digital/microsoft-yahoo-search-market-share-grows-google-s-dips/301934/ [https://perma.cc/S7B7-2HU6] (explaining that Google remains dominant in the market for search engines, holding 64% of the market share, while the next largest search engine, Bing, holds 21%); Parker Thomas, *Google's Search Engine Market Share Fell to 64%*, YAHOO! FIN. (Dec. 29, 2015), http://finance.yahoo.com/news/google-search-engine-market-share-170640710.html [https://perma.cc/YZG2-6BXB].

[117] *See* Section II.A.2; Geoffrey A. Manne & Joshua D. Wright, *If Search Neutrality Is the Answer, What's the Question?*, 2012 COLUM. BUS. L. REV. 151, 237 ("While PageRank's original algorithm is patent protected, trade secret law protects all subsequent adjustments Google makes to the algorithm.").

[118] *See* Manne & Wright, *supra* note 117, at 237.

[119] Burk, *supra* note 37, at 245 n.63; *see also* Mike Masnick, *Google Doesn't Rely on Intellectual Property for Its Leadership Position*, TECHDIRT (Nov. 12, 2009, 10:11 AM), https://www.techdirt.com/articles/20091110/0843176877.shtml [https://perma.cc/7RSD-LJN5].

Second, even though Myriad and Google may primarily rely upon "network effects"[120]—namely, the benefits of having many users—to draw in users today, at some early point in time, a data-generating invention has no users. In these early stages, a data-generating patent may play a fundamental role in allowing its holder to exclude competitors from practicing the underlying invention, thereby accelerating the aggregation of users and associated data, leading to downstream network effects.

Indeed, although Google may not view its patents as particularly valuable today, this was not the case early on in its development. Google's co-founder, Larry Page, was originally a graduate student at Stanford University when he invented the PageRank search technology.[121] In view of his student status, Page ceded all patent rights in PageRank to Stanford prior to cofounding Google.[122] Page nonetheless filed for a patent, with Stanford as the assignee, which—along with software copyrights—Google promptly licensed, even before it had a domain name or a business plan.[123]

---

[120] Network effects increase the value of a good or service to a user depending on the number of users engaging in similar acts. *See, e.g.*, Michael L. Katz & Carl Shapiro, *Network Externalities, Competition, and Compatibility*, 75 AM. ECON. REV. 424, 424–26 (1985); S.J. Liebowitz & Stephen E. Margolis, *Network Externality: An Uncommon Tragedy*, 8 J. ECON. PERSP. 133 (1994).

[121] JOHN MACCORMICK, NINE ALGORITHMS THAT CHANGED THE FUTURE: THE INGENIOUS IDEAS THAT DRIVE TODAY'S COMPUTERS 24–25 (2012). Some accounts mention Google's other co-founder, Sergey Brin, as an originator of the PageRank method. *See* Letter from Mark Fuchs, Chief Accountant, Google, Inc., to Sec. & Exch. Comm'n (Aug. 11, 2006), https://www.sec.gov/Archives/edgar/data/1288776/000119312506170952/filename1.htm [https://perma.cc/Q8FA-DAWU] ("Larry and Sergey helped create the PageRank patent."); *see also* Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 30 COMPUTER NETWORKS & ISDN SYS. 107 (1998). However, Page is the only inventor listed on the original PageRank patent. U.S. Patent No. 6,285,999 (filed Jan. 9, 1998) (issued Sept. 4, 2001).

[122] CORONA BREZINA, SERGEY BRIN, LARRY PAGE, ERIC SCHMIDT, AND GOOGLE 30 (2012).

[123] *See* U.S. Patent No. 6,285,999 (disclosing the method for Google's "PageRank" search technology); David Pridham & Brad Sheafe, *Using IP to Benefit Startups and Large Companies Alike*, CORP. COUNSEL (Aug. 25, 2015), http://www.corpcounsel.com/id=1202735624108/Using-IP-to-Benefit-Startups-and-Large-Companies-Alike-?slreturn=20151012124727 [https://perma.cc/TNQ5-4AMA] ("Google's original page rank [sic] patent [was] filed before the search giant even had a business plan or a domain name, and [was] so valuable to Google it paid Stanford $336 million in shares for an exclusive license to it."); Florian Mueller, *The 4 Big Problems with Google's Anti-Patent Stance*, BUS. INSIDER AUSTL. (July 16, 2011), http://www.businessinsider.com.au/the-4-big-problems-with-googles-anti-patent-stance-2011-7 [https://perma.cc/Q8K7-LXV3] ("[Google] had a software patent before they had a business plan, and there's at least a strong possibility that the patent played a key role in Google's ability to attract funding."); David Pridham, *"Troll" Reform Would Kill Startups*, PROVIDENCE J. (May 27, 2015, 2:01 AM), http://www.providencejournal.com/article/20150527/OPINION/150529433 [https://perma.cc/93B3-HF2S] ("When Google was a startup, after all, it filed its seminal PageRank patent before the firm even had a business plan, venture funding, or a domain name — and then paid Stanford University $336 million in shares for an exclusive license to it."); David Pridham, *How to Kill the Next Generation of Startups*, SAN JOSE MERCURY NEWS (May 19, 2015, 4:36 AM), http://www.mercurynews.com/opinion/ci_28146187/david-pridham-how-kill-next-generation-startups [https://perma.cc/TE4S-AAQL].

The original Stanford–Google license included an undisclosed up-front payment, yearly royalties, and a portion of the company's equity.[124] In 2003, Google renewed its by-then exclusive license for its original patent, along with two other undisclosed patents, for additional yearly royalties through 2011 totaling over $1 million.[125] Ultimately, Stanford sold its equity received from the deal for over $300 million, placing it in the top three licensing deals in the University's history—between the process of creating functional antibodies ($486 million) and the process of creating recombinant DNA ($255 million).[126]

If Google and its lawyers believed that its patents were of no value, then the company would not have paid sizable royalties, which constituted a large portion of its equity, to license the original patent.[127] Although this is especially so initially, when Google was arguably economizing on expenses, this was even the case in 2003, when Google was already dominant in the search space, yet renewed its exclusive license.[128]

---

[124] The original license was executed on a nonexclusive basis in 1998, but with an option to make the license exclusive, which Google exercised in 2001. *See* Google, Inc., Exclusive License (Exhibit 10.10.1 to S-1/A Filing) (Aug. 18, 2004), *available in* Lexis-Nexis EDGAR database (disclosing terms of Google's exercise of exclusivity option under original 1998 license); Google, Inc., Amended & Restated License Agreement (Exhibit 10.10.3 to S-1/A Filing) (Oct. 13, 2003), https://www.sec.gov/Archives/edgar/data/1288776/000119312504105564/dex1010.htm [https://perma.cc/4HD2-UZJL] (providing a redacted version of the 2003 license agreement between Stanford and Google).

[125] *See* Amended & Restated License Agreement, *supra* note 124. After 2011, further royalties were to be paid until the end of the patent term, with an apparent reduction in yearly royalties from the end of the patent term to the end of the copyright term of Google's original software. *See id.* The royalty amounts for 2003 to 2005 were $100,000 annually and from 2006 to 2010, $150,000 annually. Google, Inc. Letter, *supra* note 121, at ¶ 35. The exact amounts for the royalties from 1999 to 2002 and after 2011 do not appear to be available. *See id.*; Exclusive License, *supra* note 124; Amended & Restated License Agreement, *supra* note 124 (redacting annual royalty payments).

[126] *See* Greta Lorge, *Closing in on Cancer*, STAN. MAG., Jan./Feb. 2016, https://alumni.stanford.edu/get/page/magazine/article/?article_id=83001 [https://perma.cc/JU4ZRWKR] (relying on FY2013 data from the Association of University Technology Managers Statistics Analysis for Tech Transfer database and the Stanford Office of Technology Licensing).

[127] *See* Colleen V. Chien, *Reforming Software Patents*, 50 HOUS. L. REV. 325, 351 (2012) ("[T]he PageRank patent, which covered a search algorithm, arguably facilitated the transfer of technology from Stanford University to Google."). Moreover, although Stanford held the copyright to the original software, presumably the software could have been easily recreated from scratch by third-party developers in a "clean room" environment using the disclosure in Google's patent. *See* Pamela Samuelson et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308, 2318 n.24 (1994) (describing how a "clean room" approach can avoid copyright infringement).

[128] Given the large number of additions and changes to Google's search techniques, presumably the exclusive license to the copyright in Google's original code from 1999 was not of much value by 2003. *See* AMY N. LANGVILLE & CARL D. MEYER, GOOGLE'S PAGERANK AND BEYOND: THE SCIENCE OF SEARCH ENGINE RANKINGS (2011).

In particular, Google's payments for the Stanford patent provide strong circumstantial evidence that exclusivity—and, hence, the ability to prevent others (like Yahoo) from using its PageRank technology—was important to Google through at least 2003.[129] Indeed, instead of attempting to replicate Google's PageRank technology and infringe the Stanford patent, Yahoo licensed it and related search technology in the early 2000s.[130] At the same time, Google did not provide Yahoo with access to its PageRank algorithms or any of its proprietary data—rather, Yahoo merely overlaid its design elements over Google's underlying search results.[131] As such, this license did not destroy Google's effective exclusivity and control over the data generated by its search technology. In this regard, the Yahoo license may have ultimately had more value to Google than to Yahoo by providing Google with even more data that it could aggregate and use to improve its underlying product, as well as to expand into secondary markets.[132] Additionally, it appears from the available evidence that Google's patents at least in part caused some of the then-leading search engines, like AltaVista and HotBot, to avoid implementing PageRank-like

[129] *See* Stuart J.H. Graham et al., *High Technology Entrepreneurs and the Patent System: Results of the 2008 Berkeley Patent Survey*, 24 BERKELEY TECH. L.J. 1255, 1317 (2009) (describing the process of startups licensing patents).

[130] Jim Hu, *Yahoo Dumps Google Search Technology*, CNET (Feb. 18, 2004, 9:56 AM), http://www.cnet.com/news/yahoo-dumps-google-search-technology/ [https://perma.cc/KW5F-7PQ3]; Jim Hu, *Yahoo Sheds Inktomi for New Search Technology*, CNET (Jan. 2, 2002, 4:43 PM), http://www.cnet.com/news/yahoo-sheds-inktomi-for-new-search-technology/ [https://perma.cc/P2KS-55R6]; Press Release, Google News, Yahoo! Selects Google as its Default Search Engine Provider (June 26, 2000), http://googlepress.blogspot.com/2000/06/yahoo-selects-google-as-its-default.html [https://perma.cc/49Y6-7KVF]. Previously, before Google launched its website, Yahoo, Excite, and Altavista had initially rejected Brin and Page's offer to license the patents for $1 million; Page attributed this to "a lack of insight at the leadership level." WALTER ISAACSON, THE INNOVATORS: HOW A GROUP OF HACKERS, GENIUSES, AND GEEKS CREATED THE DIGITAL REVOLUTION 462 (2014).

[131] *See* Hu, *Yahoo Dumps Google Search Technology*, *supra* note 130 ("This arrangement, first viewed as a placeholder for Yahoo, eventually gave Google more exposure to Web surfers around the world, helping it become a brand synonymous with search.").

[132] In 2004, Yahoo broke with Google and decided to launch its own search engine, which of course never was able to compete effectively with Google's. *See* Hu, *Yahoo Dumps Google Search Technology*, *supra* note 130. Indeed, in 2008, Yahoo's attempts to rely on Google's search ads were blocked by the Department of Justice. *See* Vindu Goel, *In Tests, Yahoo Uses Google to Power Search Results and Ads*, N.Y. TIMES (July 1, 2015, 9:13 PM), http://bits.blogs.nytimes.com/2015/07/01/in-tests-yahoo-uses-google-to-power-search-results-and-ads/?_r=0 [https://perma.cc/UQB4-WBPW]. There is some indication that Yahoo recently returned to relying on Google's search results and search ads. *See id.* Even more recently, Yahoo sold its search engine business to Verizon, which plans to merge it into AOL—pending government approval—likely spelling further changes to Yahoo's search engine technology. *See* Kara Swisher, *As Expected, Verizon Says It Will Buy Yahoo for $4.83 Billion*, RECODE (July 25, 2016, 7:13 AM), http://www.recode.net/2016/7/25/12269882/as-expected-verizon-says-it-will-buy-yahoo-for-4-83-billion [https://perma.cc/VM6J-72VA].

technology, ultimately leading to their demise.[133] As one of the leading patent law scholars, John F. Duffy, has remarked, "the patent on Google's PageRank technology . . . is surely one of the most famous and valuable of all modern software patents."[134]

Moreover, foundational patents often serve as signaling mechanisms, making startups, such as Google at that time, more attractive to investors and ensuring their viability to compete in somewhat crowded markets.[135] Google's patented technology represented a major advance in providing quality search results, paving the way for its dominance in the search industry.[136] To be certain, Google's patents do not wholly foreclose competition in the relevant market sector (i.e., internet search), and neither Google nor Stanford have publicly enforced the PageRank patents.[137]

Additionally, much of Google's market dominance is afforded by network effects and trade secrecy.[138] Nonetheless, Google's original PageRank patent is fairly broad in scope, arguably foreclosing competitors from implementing similar "web linking" search algorithms.[139] In this

---

[133] *See* Daniel Eran Dilger, *Google's Current Stance on Patents with Android Would Have Prevented Google from Ever Having Existed*, APPLEINSIDER (May 25, 2014, 8:39 PM), http://appleinsider.com/articles/14/05/25/googles-current-stance-on-patents-with-android-would-have-prevented-google-from-ever-having-existed [https://perma.cc/9P7X-ECSS] ("If it weren't for . . . Google's exclusive rights to practice the PageRank invention, the work that Page and Brin had orchestrated and found private investment for would have been appropriated and copied by larger, better-funded existing search firms . . . ."); Mueller, *supra* note 123 ("The broad PageRank patent may also have helped deter competitors from matching Google's quality especially in its critical early years."); LANGVILLE & MEYER, *supra* note 128, at 18 (noting the benefits of the PageRank technology over competing approaches).

[134] John F. Duffy, *The Death of Google's Patents?*, PATENTLY-O PAT. L.J. 2, 3–7 (July 21, 2008), http://patentlyo.com/media/docs/2008/07/googlepatents101.pdf [https://perma.cc/485Q-8VKU].

[135] *See* Sichelman & Graham, *supra* note 29, at 123 (discussing how the signaling function may provide the primary value of a foundational patent, particularly for software startups.).

[136] *See* Manish Agarwal & David K. Round, *The Emergence of Global Search Engines: Trends in History and Competition*, 7 COMPETITION POL'Y INT'L 115, 126–27 (2011) ("Google's rise as a leading search provider demonstrates how a search engine can outperform its competitors based on superior innovation. Its search algorithm . . . brought a significant increase in the quality of search results."); Kristine Laudadio Devine, *Preserving Competition in Multi-Sided Innovative Markets: How Do You Solve a Problem Like Google?*, 10 N.C. J.L. & TECH. 59, 75 (2008); Greg Lastowka, *Google's Law*, 73 BROOK. L. REV. 1327, 1337 (2008).

[137] On the other hand, it is very likely Google has relied upon its search patents for negotiating leverage in cross-licensing deals and it is possible that it has privately threatened suit. *See* Sichelman & Graham, *supra* note 29, at 112–13 (explaining how patents can be used to generate bargaining leverage).

[138] *See* Newman, *supra* note 13, at 421–22 (discussing the benefits flowing from Google's network effects); Manne & Wright, *supra* note 117, at 237 (discussing the advantages of Google's trade secrets).

[139] *See, e.g.*, U.S. Patent No. 6,285,999 (filed Jan. 9, 1998) (issued Sept. 4, 2001) (disclosing broad claims covering, for example, a "computer implemented method of scoring a plurality of linked documents"); Duffy, *supra* note 134, at 1–2 ("Google's PageRank technology . . . is surely one of the most famous and valuable of all modern software patents.").

regard, Google's current use of its patents for so-called defensive purposes is not an act without market effect; indeed, cross-licensing agreements among incumbents in a market can afford them market power that may exclude entrants from the market.[140]

Given the apparent strength of this patent and Google's keen interest to license it from Stanford originally and again in 2003, having exercised the option to make the license exclusive, it seems reasonable to assume that other search companies chose—at least early on—not to implement Google's valuable technology out of fear of a lawsuit.[141] Like Myriad, which also relied heavily on trade secrecy and network effects to gain market power,[142] it is likely that Google has been able to leverage its patented search technology to obtain a nontrivial competitive advantage in gathering data.[143] Google captures searches, location and network data, and other information about its users.[144] And much like Myriad's use of patient information, Google's use of this proprietary information long after its patents expire or are invalidated affords it the ability not only to improve its responses to search queries, but also to offer superior targeted advertising of its users to business partners.[145]

3. Data-Generating Patents Across Technological Fields.—Data-generating patents appear in numerous other applications and fields. In the information technology area, Google and Facebook have obtained patent protection over methods of facial recognition.[146] These companies can link vast data about their users, such as search queries, locations, employment,

---

[140] *See* Roger B. Andewelt, *Analysis of Patent Pools Under the Antitrust Laws*, 53 ANTITRUST L.J. 611, 617–19 (1984) (describing various anticompetitive effects of patent pools); *see also* U.S. DEP'T OF JUSTICE & FED. TRADE COMM'N, ANTITRUST GUIDELINES FOR THE LICENSING OF INTELLECTUAL PROPERTY 30 (1995) (noting that cross-licensing agreements may "substantially reduce . . . incentives to engage in research and development and thereby limit rivalry in innovation markets").

[141] *See supra* notes 128–40.

[142] *See supra* note 120 and accompanying text.

[143] *See supra* Section II.A.1.

[144] *Privacy Policy*, *supra* note 5 (explaining that Google collects information both volunteered by users and received from their variety of services); *see also* Dominic Rushe, *Google: Don't Expect Privacy When Sending to Gmail*, GUARDIAN (Aug. 15, 2013, 3:00 AM), http://www.theguardian.com/technology/2013/aug/14/google-gmail-users-privacy-email-lawsuit [https://perma.cc/YN7K-QNXS] (stating that Google uses information obtained through its users' email to target advertising).

[145] *Terms of Service*, *supra* note 13 (describing Google's use of information for targeted advertising); *see* Grimmelmann, *supra* note 13, at 942–43 (describing how Google stores and uses search information to engage in targeted advertising); Nathan Newman, *supra* note 13 (explaining that Google is able to "extract ever more precise information about users to allow advertisers to more effectively target particular ads to those users").

[146] *See, e.g.*, U.S. Patent Nos. 8,213,689 (issued July 3, 2012); 8,437,500 (issued May 7, 2013); 8,442,265 (issued May 14, 2014).

and social networking data with their images identified through facial recognition technology.[147] They can retain and analyze this information as a trade secret long after their patents expire or are invalidated, providing vast amounts of data for targeted advertising and new technologies.[148] For instance, one potential application is placing cameras in stores to identify shoppers, and then providing personal information to salespersons about the shoppers' preferences.[149]

As the ability to aggregate and analyze data from the use of patented technology increases, various patent holders will have the capacity to maintain their user data as trade secrets. Advances in DNA analysis and genome editing will allow for collection of a wide variety of information about patients.[150] In the area of smart medical devices, such as heart monitors and fitness trackers, patent holders can gather data about locations, times, and biometric data.[151] Such devices can collect and store information about a user's actions at a given time, such as sleep habits, blood pressure, dietary intake, activity, stress levels, and productivity.[152] For example, a patent covering a smart contact lens allows for measuring glucose levels in tear fluid to help in diabetes management, as well as capturing the concentration of other substances in tear fluid.[153] Patent

---

[147] At this point, Google has stated that it will not allow facial recognition apps for Google Glass. *See* Natasha Singer, *Never Forgetting a Face*, N.Y. TIMES (May 17, 2014) http://www.nytimes.com/2014/05/18/technology/never-forgetting-a-face.html?_r=0 [https://perma.cc/P2RJ-YXAX]. The major concern with "faceprinting," or facial recognition software, is that it is being used to identify someone by name, and then the software connects the name with personal details, such as "their home addresses, dating preferences, employment," search histories, and location. *Id.* The software does all this "without [the person's] awareness or permission." *Id.*

[148] *See id.*

[149] *See* Joshua A.T. Fairfield, *Mixed Reality: How the Laws of Virtual Worlds Govern Everyday Life*, 27 BERKELEY TECH. L.J. 55, 81 (2012) (discussing the example of "a PerC-enabled mall, in which RFID chips embedded throughout the mall communicate to the shopper"); Jerry Kang & Dana Cuff, *Pervasive Computing: Embedding the Public Sphere,* 62 WASH. & LEE L. REV. 93, 101, 124 (2005).

[150] *See, e.g.*, Heidi Ledford, *Bitter Fight over CRISPR Patent Heats Up*, NATURE (Jan. 12, 2016), http://www.nature.com/news/bitter-fight-over-crispr-patent-heats-up-1.17961 [https://perma.cc/F3L9-2L7L] (describing the CRISPR patent dispute); Kelly Severick, *Amid Patent Lawsuit, Genetic Sequencing Upstart Unveils New Technology*, SCIENCE (Mar. 8, 2016, 4:15 PM), http://www.sciencemag.org/news/2016/03/amid-patent-lawsuit-genetic-sequencing-upstart-unveils-new-technology [https://perma.cc/TL6B-TAFT] (describing nanopore sequencing technology).

[151] *See* Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent*, 93 TEX. L. REV. 85, 88 (2014) (describing new medical and fitness tracking devices that "measure, record, and analyze different aspects of daily life," including fitness bands that will "track the steps you take in a day, calories burned, and minutes asleep;" bands that will track heart rate, blood pressure, and glucose levels; bands that track sports performance; and bands that track mental state).

[152] *See id.*

[153] *See, e.g.*, U.S. Patent No. 8,985,763 (filed Sept. 26, 2012) (issued Mar. 24, 2015) (covering a contact lens containing a chip); Andrew Morse, *Novartis and Google to Work on Smart Contact Lenses*,

holders and licensees of cellular phones and related mobile software can retain similar data[154]—likewise, security systems and televisions are now regularly tracking and storing user data.[155] Gaming machines are another area in which user activity, betting patterns, and other information can be tracked and maintained as a trade secret.[156] Even patent holders with protection relating to voting machines may be using the patented technology to obtain data in ways that can be protected under trade secret law.[157] Last, there is a wide class of traditional research tools—including microscopes, telescopes, spectrometers, photometers, and hundreds of other measuring instruments—that collect (and now store) data and are often patented.[158]

While the collection of information by corporate and other entities from their patented products and services is not new, advances in the means of aggregating and processing in the era of big data have dramatically altered the ability to use information.[159] For instance, utilities are investing

WALL ST. J., (July 15, 2014, 11:51 AM), http://www.wsj.com/articles/novatis-google-to-work-on-smart-contact-lenses-1405417127 [https://perma.cc/4VBD-LRTG].

[154] *See* Steven I. Friedland, *Cell Phone Searches in a Digital World: Incorporating Function as well as Form in Fourth Amendment Analysis*, 19 TEX. J. C.L. & C.R. 217, 225 (2014) (stating that cell phones and their providers have the ability to create and store huge amounts of data, including "the location of cell phones" through cell phone towers); Peppet, *supra* note 151, at 115–16 (stating that smartphones can also track the owner's health, demographics, sleep patterns, stress levels, and fitness through the vast amounts of applications available to download).

[155] *See, e.g.*, *ADT Security - Privacy Policy*, ADT, http://www.adt.com/about-adt/legal/privacy-policy [https://perma.cc/5C94-ARGQ] (last modified Aug. 18, 2015) (describing information collected); *Privacy Policy,* APPLE, http://www.apple.com/privacy/privacy-policy/ [https://perma.cc/TSG9-RVQ6] (last modified May 31, 2016) (describing information collection associated with Apple TV and other services); Claire Cain Miller, *Tech Companies, Bristling, Concede to Federal Surveillance Program*, N.Y. TIMES (June 7, 2013), http://www.nytimes.com/2013/06/08/technology/tech-companies-bristling-concede-to-government-surveillance-efforts.html?_r=0 [https://perma.cc/6SBD-4FJH] (describing how companies "were legally required to share the data under the Foreign Intelligence Surveillance Act"); Michael Lewis, *Boom Box*, N.Y. TIMES MAG., Aug. 13, 2000, at 36 (discussing how TiVo collects personal data by means similar to the internet); A. Michael Froomkin, *Flood Control on the Information Ocean: Living with Anonymity, Digital Cash, and Distributed Databases*, 15 J.L. & COM. 395, 480–81 (1996) (describing the concerns with tracking "citizens' movements" with security cameras and other technology).

[156] *See, e.g.*, Bally Techs., Inc. v. Bus. Intelligence Sys. Sols., No. 2:10-CV-00440-PMP-GWF, 2011 U.S. Dist. LEXIS 75887, at *3 (D. Nev. July 11, 2011) ("[C]asinos gather substantial amounts of data from their gaming devices, such as how much a particular machine is played, at what denomination, during what time of day or what day of the week.").

[157] *See, e.g.*, Matthew Fisher, *Will Your Vote Count?: Can the Current Software Withstand and Guarantee the Constitutional Right to Vote?*, 8 J. HIGH TECH. L. 91, 96 (2008).

[158] *See generally Scientific Instruments*, PHRONTISTERY, http://phrontistery.info/instrum.html [https://perma.cc/VGR8-97B3] (listing hundreds of scientific measuring instruments).

[159] *See supra* note 89; Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1939 (2013) (discussing processing and analyzing big data); Brenda M. Simon, *The Implications of*

in smart grids and meters that allow for the collection and analysis of vast amounts of data.[160] In addition to the standard usage of data analysis to ensure reliability in services and to minimize the length of outages, smart meter data mining can be useful in far-reaching areas, such as setting prices and marketing related products.[161]

These are but a few examples of how data mining and analysis in the era of big data, when coupled with the market exclusivity afforded by patents, can produce unique legal and economic consequences. As we explain further below, although data-generating patents certainly do not comprise the bulk of existing patents today, they are numerous and rapidly growing, and their implications for competition, public health, and privacy have a greater substantive effect than their volume reflects.

B. Legal Consequences and Concerns of Patenting Data-Generating Inventions

The traditional safeguards of patent and trade secret law sometimes do not apply to data-generating inventions. Limitations in the disclosure requirements of patent law allow for much of the information obtained from data-generating patents to remain secret long after the leveraged patents expire or are invalidated.[162] Further, the inability of would-be competitors to reverse engineer or engage in independent discovery while the underlying patent is in effect might also upset the balance between patent and trade secret law. Indeed, we suggest that the recent exclusion of some data-generating inventions from patent eligibility may stem from concerns about the broad effects of these inventions not only on downstream technologies but also on downstream data.[163]

1. Data-Generating Patents and Trade Secrets as Complements.—Data-generating patents present a unique use of

Technological Advancement for Obviousness, 19 MICH. TELECOMM. & TECH. L. REV. 331 (2013) (describing advances in big data techniques).

160 *See Utilities and Big Data: Accelerating the Drive to Value*, ORACLE 17 (July 23, 2013), http://www.oracle.com/us/dm/oracle-utilities-2013-1979214.pdf [https://perma.cc/8UUR-BE82] ("The average utility with more than one million customers will invest approximately $180 million in the smart grid and smart metering over the next five years.").

161 *Id.*

162 *See supra* note 33. There are structural similarities between our analysis and how companies can leverage their patents to enjoy greater *trademark* protection, and the responses of courts to limit such leverage. *See* Parchomovsky & Siegelman, *supra* note 31, at 1462–63 (describing how trademark and patent law may serve as economic complements so as to effectively extend patent term); *see also* Kellogg Co. v. Nat'l Biscuit Co., 305 U.S. 111, 118 (1938).

163 *See infra* Section III.B.1.

patents and trade secrets as economic complements.[164] Information can be withheld from a patent in two ways: (1) protecting innovative information as a secret prior to patenting and (2) withholding or maintaining secret information from a patent application or patent in view of fairly weak disclosure requirements. Data-generating patents operate along the second dimension, particularly after the issuance of a patent.[165]

In this regard, once a patent is filed, there is no duty to update its disclosure with newly found information relevant to the patent; indeed, there is an affirmative proscription against doing so.[166] As such, inventors may choose to retain post-filing information about the invention as a trade secret. Typically, such information may include ideal operating parameters, improvements, and general market information about the invention. In this regard, commercial embodiments of inventions developed well after patenting often substantially differ from what is disclosed in a patent.[167] Yet, because the scope of the legal rights afforded by patents is generally much broader than what is disclosed in the patent disclosure, these commercial embodiments—and the information related to them—typically are protected by the original patent or by trade secret law.[168]

For instance, an inventor may, after filing for a patent, determine the ideal material for a collapsible garden hose, whose novelty mainly depends on its unique shape.[169] Because the inventor did not know of the material at the time of filing, there was no need to disclose it then (*or at any future time*), and if it is difficult to reverse engineer, the inventor may effectively retain it as a trade secret.[170] Because the patent provides the inventor the ability to exclude others from manufacturing any hose of the same or similar shape, there will be little incentive (absent a license from the patent holder) for others to perform research to determine the ideal material for the hose during the term of the patent.[171] Indeed, making such a hose, even if purely for research purposes, would constitute infringement.[172] Once the

---

[164] *See supra* Section I.B.

[165] *See supra* notes 72–80 and accompanying text.

[166] 35 U.S.C. § 132 (2012) (barring the addition of "new matter" to a patent application so as to ensure all material in the application originates from the priority date or earlier).

[167] *See supra* note 73.

[168] *See supra* notes 76–80.

[169] *See, e.g.*, U.S. Patent Application No. 13/690,670 (filed Nov. 30, 2012) (disclosing an "expandable garden hose").

[170] *See supra* notes 76–80.

[171] § 271; Roger D. Blair & Thomas F. Cotter, *Strict Liability and Its Alternatives in Patent Law*, 17 BERKELEY TECH. L.J. 799, 800 (2002) ("Patent infringement is often characterized as a strict liability tort . . . .").

[172] § 271; *infra* note 345.

term expires (or perhaps right near the end if detection is unlikely), a third party may begin to experiment with different materials to create a "generic" copy of the hose.[173] The ability of the original patentee to maintain the material as a trade secret, however, will allow it to effectively extend its patent rights past the twenty-year term afforded by the patent, delaying third-party attempts to determine the ideal material after expiration.[174] In other words, if potential competitors cannot determine the material, or a close substitute, the patentee will continue to be able to charge above-market prices for its now *un*patented hose.[175]

To the extent inventions are supposed to fall into the public domain once the patent term expires, this use of patents and trade secrets as complements arguably thwarts the goals of the patent system.[176] Even in fields in which some disclosure of post-patenting data is required for regulatory purposes, such as medical devices or pharmaceuticals, this strategy can still be effective because such disclosed data is typically maintained as a secret by the applicable regulatory agency.[177] Indeed, even third-party testing of medical-related inventions to determine efficacy and safety will generally constitute actionable infringement.[178]

These limitations of the current patent disclosure rules have been well studied.[179] Data-generating patents, however, present unique concerns that have not been sufficiently addressed. Unlike the garden hose invention, a data-generating patent will often yield information subject to trade secret law not merely about the invention itself, but also about the invention's users and that can be advantageous in separate product markets.[180] As noted, Google's patented search algorithms can be used not only to

[173] Julie E. Cohen & Mark A. Lemley, *Patent Scope and Innovation in the Software Industry*, 89 CALIF. L. REV. 1, 24 n.86 (2001) ("One classic justification for having a patent system is to encourage inventors to disclose their ideas to the public, who will benefit from this new knowledge once the patent expires."); Simon, *supra* note 99, at 1342 (2011) ("Most research use is ignored until it is disseminated or results in an improvement or alternative means.").

[174] Cohen & Lemley, *supra* note 173.

[175] *Id*.

[176] *See supra* note 45.

[177] *See infra* note 358 and accompanying text; *see also* W. Nicholson Price II & Arti K. Rai, *Manufacturing Barriers to Biologics Competition and Innovation*, 101 IOWA L. REV. 1023, 1028 (2016) (discussing trade secret protection for manufacturing processes of patented biologics); Price II, *supra* note 58, at 32 (describing how innovators receive both patent protection on biologics and "are sheltered from competition indefinitely by the secrecy surrounding their manufacturing methods").

[178] Simon, *supra* note 99, at 1324–25; *cf.* 35 U.S.C. § 271(e) (2012) (providing for an exception to infringement for experimentation related to the submission of information to the FDA).

[179] *See* Sichelman, *supra* note 76, at 355–56; Jeanne C. Fromer, *Patent Disclosure*, 94 IOWA L. REV. 539, 572–73 (2009).

[180] *See supra* Section II.A.

improve those algorithms but also to better target advertising to its users.[181] Such leveraging into secondary markets in the era of big data presents a use of patents and trade secrets as economic complements that was essentially unheard of in the twentieth century and arguably well outside the "bargain" for exclusive rights contemplated by the patent system.

2. *Traditional Trade Secret Safeguards Generally Do Not Apply to Data-Generating Patents*.—Ordinarily, trade secrets are subject to reasonable limitations to prevent against overreaching.[182] Two oft-cited defenses in ensuring that trade secret law does not overreach are the ability to reverse engineer a product or process embodying a trade secret, or simply independently discovering the trade secret.[183] In contrast, neither reverse engineering nor independent discovery typically provide defenses to patent infringement.[184] Indeed, the Supreme Court in *Kewanee Oil Co. v. Bicron Corp.* focused on independent creation and reverse engineering as the key distinguishing features in determining that trade secret was not preempted by patent law.[185] If either occurs, trade secret protection will be lost.[186]

With data-generating patents, the ability to reverse engineer the resulting trade secret is often foreclosed, or at least severely restricted, during the term of the data-generating patent. Traditionally, competitors and others can obtain information about trade secrets by observation or through analysis of products or other materials that are publicly accessible.[187] Reverse engineering has a major benefit: it "will often generate knowledge about the product being reverse engineered that will make it possible to improve on it."[188] As David Friedman, William Landes, and Richard Posner have suggested, without the ability to effectively reverse engineer, trade secrets would become "superpatents" that would have "obvious problems of conflict with patent law."[189] The major difference with patenting is twofold: (1) patents have a limited duration

181 *See supra* note 145 and accompanying text.

182 *See* Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 489–90 (1974).

183 *Id.* at 476; *see also* Anderson, *supra* note 41, at 950.

184 *See Kewanee Oil Co.*, 416 U.S. at 489–90.

185 *Id.* at 490 ("[T]rade secret law does not forbid the discovery of the trade secret by fair and honest means, *e.g.*, independent creation or reverse engineering . . . .").

186 *See id.*; Jon Chally, Note, *The Law of Trade Secrets: Toward a More Efficient Approach*, 57 VAND. L. REV. 1269, 1285 (2004).

187 UNIF. TRADE SECRETS ACT § 1 cmt. (UNIF. LAW COMM'N 1985); *Kewanee Oil Co.,* 416 U.S. at 476 (defining reverse engineering).

188 *See* Friedman, Landes & Posner, *supra* note 56, at 70.

189 *Id*. at 70–71 ("The fact that reverse engineering is costly, moreover, automatically cuts down on the amount of free-riding on the first inventor.").

while trade secrets do not and (2) patents by definition disclose valuable information about the innovation.[190] Trade secret protection is designed to expire at the point the secret can be reversed engineered or independently invented—such a balance makes economic sense because society loses little by enforcing a trade secret that others could not have generated on their own.[191]

In the case of data-generating patents, however, the ability to reverse engineer trade secrets is often illusory.[192] When a data-generating invention is patented, would-be competitors are effectively foreclosed from reverse engineering for a substantial period, as they are prohibited from making and using the patented invention during the twenty-year exclusivity period.[193] Whether the underlying data-generating invention is a patented search engine or medical diagnostic test, the ability to reverse engineer or independently generate otherwise secret information obtained through patent exclusivity often is infeasible.[194] In particular, a search engine or

[190] *See* Beckerman-Rodau, *supra* note 33, at 383, 394; *see also* 35 U.S.C. § 112 (2012).

[191] *See* Rochelle Cooper Dreyfuss, *Trade Secrets: How Should We Be Allowed to Hide Them? The Economic Espionage Act of 1996*, 9 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1, 16 (1998) (explaining that without reverse engineering, trade secret protection would essentially last forever and inventors would select trade secret over patent).

[192] *See Kewanee Oil Co.*, 416 U.S. at 490 (1974) ("While trade secret law does not forbid the discovery of the trade secret by fair and honest means, *e.g.,* independent creation or reverse engineering, patent law operates 'against the world,' forbidding any use of the invention for whatever purpose for a significant length of time.").

[193] *See id.*; *see also* Fauver, *supra* note 9, at 676 (explaining that original patent holders would have established head start in the market, which puts the patent holder at an "advantage over competitors having to start 'from scratch' to develop the product").

[194] Moreover, in the case of medical diagnostic tests, even if the patent covering the data-generating invention at issue has expired, laws such as the Health Insurance Portability and Accountability Act (HIPAA) and the Genetic Information Nondiscrimination Act (GINA) often limit the distribution of the collected information. H.I.P.A.A. Privacy Rule, 67 Fed. Reg. 53,182 (Aug. 14, 2002) (to be codified at 45 C.F.R. pt. 160 and 164), https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/privacyrule/privrulepd.pdf [https://perma.cc/MTL8-489D]; Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (2008).

However, companies that develop hardware or software programs that aggregate and analyze medical data are not expressly subject to HIPAA. *See id; see also* Jeff Rabkin & Jessica Jardine Wilkes, *Wearable Technology, Big Data and the Legal Frontier*, LAW.COM (Dec. 14, 2015), http://www.law.com/sites/lawcomcontrib/2015/12/14/wearable-technology-big-data-and-the-legal-frontier/ [https://perma.cc/D5KG-PRD3]. In addition, devices that collect and analyze data without controlling the function of a medical device are generally exempt from FDA oversight, unless their use could interfere with clinical treatment. FOOD & DRUG ADMIN., U.S. DEP'T OF HEALTH AND HUMAN SERVS. ET AL., MEDICAL DEVICE DATA SYSTEMS, MEDICAL IMAGE STORAGE DEVICES, AND MEDICAL IMAGE COMMUNICATIONS DEVICES 5–8 (Feb. 9, 2015), http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM401996.pdf?source=govdelivery&utm_medium=email&utm_source=govdelivery [https://perma.cc/9XWF-FXFD].

diagnostic test patent may entirely preclude substitute technologies, which in turn precludes competitors and others from generating similar data.[195]

As noted earlier, using its patent on the BRCA gene markers for breast cancer, Myriad has been able to amass a database of mutations unrivaled by competitors or public actors.[196] Indeed, Myriad effectively stopped contributing to public databases almost a decade ago, when it decided to maintain its users' data as a trade secret.[197] Although Myriad has published articles on its findings, it has not provided its interpretive algorithms or data supporting its conclusions.[198] Any company that wants to compete with Myriad can only interpret variations based on limited public data using incomplete analytic algorithms.[199] By leveraging its patents, Myriad has managed to extend its exclusivity, even after its patents were invalidated by the Supreme Court for lack of patentable subject matter.[200] What began with patent protection over genetic information now includes trade secret protection for Myriad's databases of patients' full genetic sequences and phenotypic information, as well as the correlations and algorithms resulting from access to that wealth of data.[201] In many ways, the generated data may prove more valuable than the now-invalidated patents on the underlying technology itself.[202]

Arguably, a similar analysis as to Myriad's database applies to Google's proprietary user data obtained through its patented search technology.[203] In any technological field in which data-generating patents provide a substantial market advantage, third parties may be hindered from reverse engineering data solely protected by trade secret law.

---

195 *See supra* Section II.A.

196 *See supra* Section II.A.1; Oliver, *supra* note 113, at 540–41 ("Myriad, through its patent-conferred role as the exclusive testing laboratory for BRCA gene testing, created a database full of VUS data . . . . Since November 25, 2004, however, Myriad has kept a tight handle on its data set, retaining its important data as proprietary.").

197 *See* Cook-Deegan et al., *supra* note 37, at 586 (showing that after 2005, Myriad decided to not disclose any of its findings from its testing and began retaining "data as a trade secret").

198 *See id*. (explaining that the public lacks access to the "analytic algorithms [and] underlying sequence data").

199 *See* Oliver, *supra* note 113, at 540–41; *see also* Conley et al., *supra* note 37, at 615 ("Myriad has access to public databases in interpreting mutations, but outsiders do not have access to Myriad's database.").

200 *See* Conley et al., *supra* note 37, at 600, 616.

201 *See* Oliver, *supra* note 113, at 548 ("The VUS data obtained from Myriad's period as the exclusive provider of BRCA gene testing will likely be eligible for trade secret protection.").

202 Indeed, some have questioned whether there is still a role for patents in the personalized medicine space after recent Supreme Court decisions limiting patent-eligible subject matter. *See, e.g.*, Burk, *supra* note 37, at 236–39 (explaining the effects, positive and negative, of the Supreme Court's *Myriad* decision on personalized medicine).

203 *See supra* Section II.A.2.

In other areas of intellectual property, Congress has addressed situations in which enforcement may frustrate reverse engineering.[204] By enacting the Digital Millennium Copyright Act (DMCA), for example, Congress bolstered protection for copyrighted works by limiting the circumvention of access controls.[205] Congress, however, permitted limited exceptions for reverse engineering to allow for interoperability, encryption research, and security testing.[206] These exceptions to the DMCA highlight the need to balance robust intellectual property protection against the traditional public safeguards ensured by allowing reverse engineering of trade secrets. Unlike the DMCA, there is no exception to patent infringement for reverse engineering trade secrets obtained through data-generating patents.[207] Thus, data-generating patents may hinder data sharing, further research, and testing in ways that may be socially detrimental.

Although not quite as difficult as reverse engineering trade secret data resulting from data-generating patents, independent discovery is also severely limited by the nature of these patents. If the patent holder is dominant in the marketplace, it can preclude others—even those who independently invent—from practicing the patented invention in order to maintain its large market share. Using non-infringing techniques to generate the same data may in many instances be infeasible. For the mutations at issue in the Myriad patent, there are no other genes that can be used to identify the mutations.[208] For Google's PageRank patents, there are clearly alternatives, but they are considered inferior and thus have garnered many fewer users.[209] When there are no equivalent alternatives to a patented data-generating invention, it may be effectively impossible to recreate the data—even fully independently—generated by the data-generating patent holder.[210]

Even after a data-generating patent expires or is invalidated, the patentee will tend to have a significant lead time that will handicap any

[204] *See, e.g.*, Chamberlain Grp., Inc. v. Skylink Techs., Inc., 381 F.3d 1178, 1197–1204 (Fed. Cir. 2004).

[205] 17 U.S.C. §§ 1201(a)(1)(A), (g) (2012).

[206] *Id.* § 1201(f), (g), (j); *see also id.* § 906 (protecting reverse engineering under the Semiconductor Chip Protection Act of 1984).

[207] 35 U.S.C. § 271(a) (2012); *see also supra* notes 184–86 and accompanying text.

[208] *See supra* notes 96–99 and accompanying text.

[209] *See supra* note 116 and accompanying text; MIKLOS SAVARY, GURUS AND ORACLES: THE MARKETING OF INFORMATION viii (2012) ("PageRank made Google the dominant Internet search engine . . . .").

[210] *See supra* notes 96–99 and accompanying text (discussing how Myriad's patents provided the exclusive ability to collect patient data for over a decade).

competitor from entering the field post-expiration or post-invalidation.[211] This informational deficiency for any new market participant may result in prohibitive switching costs for consumers, even if the competitor has better methods or lower pricing.[212] The high costs of independently generating the data that the patentee obtained with the twenty-year lead time often ensures there will be little risk of independent discovery, and ultimately, competition.[213]

3. *Is the Subject Matter of Data-Generating Patents Outside the Scope of Patent Protection?*—Recent decisions from the Supreme Court have, to some degree, made it more difficult to obtain data-generating patents, particularly in the fields of medical diagnostic testing and software.[214] Most importantly, in *Mayo Collaborative Services v. Prometheus Laboratories, Inc.*, the Court expressed a concern that patents over laws of nature "will inhibit future innovation premised upon them, a danger that becomes acute when a patented process amounts to no more than an instruction to 'apply the natural law.'"[215] Based upon this rationale, the Court not only invalidated the medical diagnostic at issue, but cast doubt over the patentability of nearly all medical diagnostic tests.[216] The

[211] *See generally* Fauver, *supra* note 9, at 676 (discussing lead time advantages generated by patents).

[212] *See generally* Paul Klemperer, *Entry Deterrence in Markets with Consumer Switching Costs*, 97 ECON. J. 99 (1987).

[213] Although it may appear that the switching costs for consumers from Google to another search engine are low because users can easily choose to use a different search engine, such an analysis is wanting. *But see* Aaron S. Edlin & Robert G. Harris, *The Role of Switching Costs in Antitrust Analysis: A Comparison of Microsoft and Google*, 15 YALE J. L. & TECH. 169, 195 (2013) ("[C]hanging the default setting to a different search engine requires only a few simple steps. For that reason, default search engine contracts with OEMs do not generate significant switching costs."). Rather, the switching costs inhere in the opportunity and other costs generated by the additional time it would take a user to wade through inferior search results when using an alternative search engine that does not have access to the user's (and other users') search histories and browsing patterns. *See* ERIC ENGE ET AL., THE ART OF SEO 625 (2012) (describing how Google personalizes search results based on a user's search history); *infra* note 232.

[214] *See* Mayo Collaborative Servs. v. Prometheus Labs., Inc., 566 U.S. 66, 77, 92 (2012) (concluding that Prometheus Laboratories's patent claims were invalid because they "effectively claim the underlying laws of nature themselves" and only "simply describe these natural relations"); Alice Corp. Pty. Ltd. v. CLS Bank Int'l, 134 S. Ct. 2347, 2352 (2014) ("[M]erely requiring generic computer implementation fails to transform that abstract idea into a patent-eligible invention."); *see also* Aria Diagnostics, Inc. v. Sequenom, Inc., 726 F. 3d 1296, 1302–04 (Fed. Cir. 2013) (holding the "method for detecting a paternally inherited nucleic acid" unpatentable as a law of nature).

[215] *Mayo Collaborative Servs.*, 566 U.S. at 86 (citing Mark A. Lemley, Michael Risch, Ted Sichelman & R. Polk Wagner, *Life After* Bilski, 63 STAN. L. REV. 1315 (2011)).

[216] *See, e.g.*, Rebecca S. Eisenberg, *Diagnostics Need Not Apply*, 21 B.U. J. SCI. & TECH. L. 256, 256 (2015) (stating that the "most important advances" in diagnostics "lie outside the boundaries of patent-eligible subject matter"); Rachel E. Sachs, *Innovation Law and Policy: Preserving the Future of Personalized Medicine*, 49 U.C. DAVIS L. REV. 1881, 1908 (2016) ("[D]iagnostic method innovators

Court adopted a similar rationale in *Alice Corp. v. CLS Bank International*, where it applied the *Mayo* test to abstract ideas, and like *Mayo*, called into question the patentability of a large class of inventions, here software.[217]

Many data-generating inventions are likely to fall into the fields of medical diagnostics and software.[218] Indeed, data-generating inventions are but one example of how inventions in these areas tend to exhibit preemptive potential and allow for possible control over unforeseen markets.[219] Thus, the Court's move to limit patent protection in these fields indirectly responds to some of the specific concerns we raise about data-generating inventions.

Yet, the indirect response of limiting patentability does not eliminate all of the potential concerns data-generating inventions raise. As an initial matter, data-generating patents that predate these decisions may have already provided a market advantage to their holders, potentially requiring a response to diminish any detrimental effects.[220] Even going forward, many classes of data-generating inventions remain patentable, including smart medical and sport devices, such as pacemakers and fitness trackers; GPS and other transportation tracking devices; mobile phones, televisions, and other electronic devices, which can monitor and store usage patterns; voting machines; and security systems.[221] Also, some classes of software and medical diagnostic tests still remain patentable.[222]

Nor is it likely that patent laws in this context somehow preempt state trade secret protection for the data generated by these inventions. As the Supreme Court held in *Kewanee Oil Co. v. Bicron Corp.*, "the extension of trade secret protection to clearly patentable inventions does not conflict

---

have begun to and will continue to find it more difficult to obtain method patent claims on their inventions.").

217 134 S. Ct. at 2353–57 ("[M]ethod claims, which merely require generic computer implementation, fail to transform that abstract idea into a patent-eligible invention . . . .").

218 *See supra* Section II.A.

219 *See infra* Section III.A.

220 *See infra* Part III.

221 *See Alice Corp.*, 134 S. Ct. at 2358 (finding that applied processes may be eligible for patent protection); *see also* J. Jonas Anderson, *Applying Patent-Eligible Subject Matter Restrictions*, 17 VAND. J. ENT. & TECH. L. 267, 269 (2015) (discussing "whether the resulting [eligibility] doctrines can be predictably applied by the institutions involved in patent law").

222 Ass'n for Molecular Pathology v. Myriad Genetics, Inc., 133 S. Ct. 2107, 2119 (2013) ("cDNA is not a 'product of nature' and is patent eligible under § 101, except insofar as very short series of DNA may have no intervening introns to remove when creating cDNA."); DDR Holdings, LLC v. Hotels.com, L.P., 773 F.3d 1245, 1258–59 (Fed. Cir. 2014) (finding that the software patent claims at issue "specify how interactions with the Internet are manipulated to yield a desired result" and "recite an invention that is not merely the routine or conventional use of the Internet").

with the patent policy of disclosure."[223] Although the Court reaffirmed in *Kimble v. Marvel Entertainment, LLC* that patent royalties could not extend past the term of the patent, it expressly found that post-term royalties premised on trade secret protection were fully allowable.[224] So while it is possible to make seemingly credible arguments that data-generating patents improperly extend the scope of the patent, it seems very unlikely that such arguments would prevail.[225] For similar reasons, defenses of patent misuse in this context are unlikely to have merit.[226]

In sum, despite substantial changes in patentable subject matter, many types of data-generating inventions remain patentable, and the information generated by these inventions very likely remains protectable by state and federal trade secret law. In the next Section, we address the economic effects of these inventions in greater detail.

C. The Economic Effects of Data-Generating Patents

The economic effects of data-generating patents may raise concerns, particularly with regard to innovation incentives and the effects on the market relationship between patent holders and consumers. Although providing patent protection for data-generating inventions helps to ensure optimal incentives exist to produce these inventions and the data generated by them, too much protection may ultimately hinder downstream innovation. Additionally, providing trade secret protection for data generated during the patent term may extend deadweight losses to consumers after the underlying patent expires or is invalidated.

---

223 416 U.S. 470, 491 (1974); *cf.* Sharon K. Sandeen, Be Careful What You Wish For: Trade Secrets and the America Invents Act 1 (Sept. 20, 2013) (unpublished manuscript), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2327263 [http://perma.cc/93R6-97SG] (examining "whether in light of the AIA certain parts of trade secret law 'as applied' are now preempted by patent law").

224 135 S. Ct. 2401, 2408 (2015).

225 Burk, *supra* note 37, at 252–53 ("Preemption of trade secrets of course could prove messy; one would not want to open the flood gates to a run of employee data pilfering or industrial espionage, prompted by the prospect of hiding behind federal supremacy.").

226 Patent misuse is a defense to infringement that occurs when the patentee engages in certain behavior considered to be impermissibly anticompetitive. *See* Mallinckrodt, Inc. v. Medipart, Inc., 976 F.2d 700, 702–04 (Fed. Cir. 1992) ("The policy purpose was to prevent a patentee from using the patent to obtain market benefit beyond that which inheres in the statutory patent right."). Dan Burk contends that misuse could be used to cabin trade secret protection as "a sort of 'fruit of the poisonous tree' exclusion for leveraging patent benefits beyond the relevant patent term." Burk, *supra* note 37, at 253. Although we agree with Burk as a conceptual matter, given the Supreme Court's broad holdings in *Kewanee Oil* and *Kimble*, we do not think courts are likely to rely upon the misuse doctrine in the context of data-generating patents. We explore the issue of whether this doctrine should, as a normative matter, be altered in Part III.

1. Data-Generating Patents and Innovation Incentives.—We explained earlier how data-generating patents can yield unique market rewards by conferring exclusive control over not only the underlying invention but also the data generated by it, often for use in secondary product markets.[227] In this discussion, we suggested that because this additional protection was not contemplated by the patent system, it could potentially overcompensate inventors for their efforts.

Another possibility, however, is that this additional protection is necessary to provide optimal incentives to create data-generating inventions, or the data generated by them.[228] On this account, data-generating inventions are unique, because their very nature is to produce data.[229] For instance, as we explained, Myriad's medical diagnostic test for detecting the incidence of breast cancer is not particularly valuable absent a large database of information about genetic mutations.[230] Perhaps such a database can only be amassed by a monopolist.[231] Similarly, the more data Google collects about its users, the more valuable its search algorithms are to its users.[232] To the extent the data can be used in secondary product markets, such as consumer-targeted advertising, that is arguably a small price to pay for the value delivered via the improved underlying invention.[233]

Indeed, one can strengthen these arguments by relying upon alternatives to the traditional "reward" theory of patents, namely prospect and commercialization theory.[234] Popular variants of these theories counsel in favor of expansive patent protection in order to prevent inefficient allocation of resources by concentrating investments in R&D into a single party.[235] These benefits might extend to data-generating patents, where the data generated improves on the patented invention, or even provides

[227] *See supra* Section II.B.1.

[228] *See* Burk, *supra* note 37, at 242–47 (describing the beneficial role patents may serve in promoting data aggregation in the context of personalized medicine).

[229] *See id.* at 248 ("What is of primary interest in the context of personalized medicine is . . . the development of unpatented but valuable information.").

[230] *See id.* at 244–45 (explaining how patents in the field of personalized medicine may allow for the aggregation of otherwise "widely dispersed" information).

[231] *See id.* at 244–47 (suggesting that either monopolization or centralization may be essential to aggregate sufficient data in the context of personalized medicine).

[232] Todd Hixon, *Down-to-Earth Value from Big Data,* FORBES (Oct. 17, 2013, 6:51 AM), http://www.forbes.com/sites/toddhixon/2013/10/17/down-to-earth-value-from-big-data/#40f981ae1077 [http://perma.cc/AXT7-35ZH].

[233] *See supra* note 145 and accompanying text.

[234] *See* Burk, *supra* note 37, at 246–47 (examining the role of prospect theory in the area of personalized medicine).

[235] *See* Kitch, *supra* note 60, at 276.

inroads to another technology or market.[236] The coordination of resources by a data-generating patent holder could mitigate downstream development costs, particularly as courts place greater restrictions on subject matter eligibility.[237] For example, when Myriad's BRCA1 and BRCA2 patents were in full effect, providing it with exclusive control over the data generated by its patents, Myriad was able to reduce the rate of variants of unknown significance (VUS), providing more complete test results for patients with these mutations.[238] If the data had been generated by multiple competitors, each company's VUS rate would be significantly higher, unless the data were shared.[239]

Such an account is subject to some doubt, however. First, although broader protection might provide greater incentives to invest in inventions that generate valuable data, such an argument could be made for many types of inventions.[240] In general, broader rights will lead to more initial invention.[241] But broader rights have substantial downsides, including hindering potential downstream invention and consumer deadweight losses, which we discuss further below.[242] Additionally, broader rights in particular technological fields can create allocative distortions, channeling too much R&D into those areas relative to others that are more deserving.[243]

Second, prospect theory has been the subject of criticism, particularly because coordination often depends on the presence of low transaction costs.[244] The main concern, particularly for data-generating patents, is that innovation and improvements may be thwarted by providing too much protection for this type of data.[245] Although patent holders will often be in

---

[236] *See* Burk, *supra* note 37, at 247–49 (discussing the use of patents to coordinate the extraction of non-patent data).

[237] *See id.*

[238] *See supra* notes 106–07 and accompanying text.

[239] As we discuss in Section III.B.2, there are substantial incentives for competitors not to share data. *See infra* note 371 and accompanying text.

[240] *See* Kitch, *supra* note 60, at 268.

[241] *See id.*

[242] *See infra* Section II.C.2.

[243] *See* Amy Kapczynski & Talha Syed, *The Continuum of Excludability and the Limits of Patents*, 122 YALE L.J. 1900, 1913–16 (2013).

[244] *See, e.g.*, Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989 (1997); Peter S. Menell, *The Property Rights Movement's Embrace of Intellectual Property: True Love or Doomed Relationship?*, 34 ECOLOGY L.Q. 713, 729 (2007) (noting that the prospect theory of "[t]he optimality of private property as the model for intellectual property protection depends . . . on low transaction costs" and that "the assumption that low transaction costs are vital does not bear out in the history of many fields of innovation").

[245] Lee Petherbridge, Comment, *Intelligent Trips Implementation: A Strategy for Countries on the Cusp of Development*, 22 U. PA. J. INT'L ECON. L. 1029, 1060 (2001) ("It is also a common premise that allowing too much protection in the form of patents is economically inefficient.").

the best position to make improvements or extend technology into new areas, they may choose not to do so, particularly where they have invested in a technology that has become the standard in a field and might be superseded by an improvement.[246] Without competition, Myriad and other holders of data-generating patents will often lack the incentives to fully develop their diagnostic tests. For example, for over a decade, Myriad failed to adopt comprehensive testing mechanisms to detect mutations that were critical to patient health.[247] Myriad only made improvements to its tests after several studies conducted by nonclinical researchers described the shortcomings of Myriad's offerings.[248]

The reluctance to improve a technology and the possibility of high transaction costs caution against merely expanding the scope of patent rights to enhance coordination, at least without some contrary empirical evidence.[249] As there is little to no evidence showing that incentives to produce data-generating inventions are too weak absent further legal protection, we believe a cautious stance towards strengthening patent protection in this instance is called for, especially given the potential social costs of these inventions.[250]

2. The Extension of Deadweight Losses by Data-Generating Patents.—Aside from innovation incentives, data-generating patents have the potential to significantly affect the market relationship between patent holders and consumers. Typically, part of the quid pro quo

---

[246] Cont'l Paper Bag Co. v. E. Paper Bag Co., 210 U.S. 405, 428–29 (1908); Merges & Nelson, *supra* note 63, at 843, 907–09; Kurt M. Saunders, *Patent Nonuse and the Role of Public Interest as a Deterrent to Technology Suppression*, 15 HARV. J.L. & TECH. 389, 423 (2002); *see also* Simon, *supra* note 99, at 1317.

[247] *See* Bryn Williams-Jones, *History of a Gene Patent: Tracing the Development and Application of Commercial BRCA Testing*, 10 HEALTH L.J. 123, 138–39 (2002); Eileen M. Kane, *Patent-Mediated Standards in Genetic Testing*, 2008 UTAH L. REV. 835, 849 ("Scientific reports that identify genomic arrangements in the BRCA1 and BRCA2 genes that are not detected by the Myriad testing date back at least to 2001."); Sachs, *supra* note 216, at 1892 n.48 ("Scientists alleged that as late as 2006, more than ten years after it began operating, Myriad's test still did not detect a significant percentage of large genomic deletions or duplications."); Simon, *supra* note 99, at 1309 ("A study by nonclinical researchers has shown that Myriad's exclusive testing provided false negatives in 10–20% of high risk patients."); Tom Walsh et al., *Spectrum of Mutations in BRCA1, BRCA2, CHEK2, and TP53 in Families at High Risk of Breast Cancer*, 295 JAMA 1379, 1380, 1386 (2006) (stating that multiple studies have identified gene mutations that were not detected by Myriad's testing methods).

[248] Simon, *supra* note 99, at 1309 n.47 ("Though Myriad eventually adopted more comprehensive methods of testing, how long might Myriad have taken to implement these methods absent the dissemination of information from the study?").

[249] *See* Sichelman, *supra* note 76, at 381–89 (explaining how in the presence of high transaction costs, expanding the scope of patent rights may decrease technology commercialization).

[250] *See* Burk, *supra* note 37, at 253–54; Conley et al., *supra* note 37, at 600; Cook-Deegan et al., *supra* note 37, at 585.

for granting patent exclusivity is the assurance that information about the invention will become part of the public domain after patent expiration.[251] By protecting information resulting from data-generating patents, however, it may be too costly for competitors to use the invention itself. Thus, while the data-generating invention technically falls into the public domain following expiration, the ability to retain key data generated by the invention may effectively preclude patent law's quid pro quo. Of course, the limitations of the disclosure requirements, such as the failure to require applicants to update their disclosures after filing, always result in some information being retained by the patent holder. What makes data-generating inventions different from other inventions is that the retained information does not just concern the use or operation of the invention itself, but rather can affect disparate markets for the information, and for a potentially indefinite duration.[252]

Patents can impose substantial consumer losses during the patent term.[253] In other words, if a patent provides market power to its holder, prices of the patented good will tend to be higher than the competitive price, which will foreclose some consumers from purchasing the patented product who otherwise would have in a competitive market.[254] These priced-out consumers create so-called deadweight losses.[255] Maintaining information contributed by consumers during the patent term as a trade secret could thus extend deadweight losses to the same class of consumers well past the expiration or invalidation of the patent. If, as we explained in the previous Section, patent term is optimally calibrated to induce innovation, then these additional deadweight losses are only worth the cost if the information protected by trade secret law provides social benefits greater than these costs.

To the extent continued deadweight losses are not socially optimal, consumers have few means to directly combat their implicit cooperation in extending these costs.[256] As an initial matter, although users are the main contributors of information in data-generating patents, they often undervalue their data.[257] Even if consumers did place greater weight on

---

[251] *See supra* note 45 and accompanying text.

[252] *See supra* notes 187–203 and accompanying text.

[253] *See* Rochelle Cooper Dreyfuss, *Are Business Method Patents Bad for Business*?, 16 SANTA CLARA COMPUTER & HIGH TECH. L.J. 263, 275 (2000) ("All patents impose social costs.").

[254] *See id.*

[255] *See id.*

[256] *See infra* notes 257–69.

[257] *See, e.g.*, Nathan Newman, *The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google*, 40 WM. MITCHELL L. REV. 850, 860–63 (2014) (stating that users "clearly underestimate the economic value of [their] personal data").

their contributions, they generally lack a property right in their data in the United States, even when they have paid to use a patented invention.[258] In addition, because data-generating inventions are protected by patents, there is often only one or a few providers of an essential service or test that participants can rely upon.[259] Thus, consumers are frequently locked in to the patent holder's business strategy, including contractual requirements to release any rights in personal data.[260]

Indeed, consumers—at least in the United States—have not been overly concerned about the use of their information by private companies, particularly when the consent process has been clear.[261] In the genetic testing space, despite complaints about Myriad's decision to discontinue data contributions to the National Institutes of Health (NIH) repository almost a decade ago, consumers (which we define broadly to include patients, genetic counselors, and physicians) still chose to use Myriad for genetic analysis.[262] Once the Supreme Court invalidated several of Myriad's patent claims, some consumers began seeking out other providers, such as Ambry and Veritas Genetics.[263] Despite the availability of recent market entrants, however, many users have nonetheless chosen Myriad, as there are still prohibitive switching costs—particularly given Myriad's decade of lead time gained largely as a result of its data-generating patents.[264]

In the area of internet search, despite repeated complaints about privacy policies, participation generally appears to have been unaffected.[265]

---

[258] Wash. Univ. v. Catalona, 490 F.3d 667, 675 (8th Cir. 2007), *cert. denied*, 552 U.S. 1166 (2008) (holding that if participants have donated samples with informed consent, they no longer hold an ownership right in them); Greenberg v. Miami Children's Hosp. Research Inst., Inc., 264 F. Supp. 2d 1064, 1074–76 (S.D. Fla. 2003) (denying parents a property right in bodily tissues and genetic information of their participant children suffering from a rare disease); Moore v. Regents of the Univ. of Cal., 793 P.2d 479 (Cal. 1990), *cert. denied*, 499 U.S. 936 (1991) (denying the plaintiff a property right in the cell line derived from his spleen).

[259] *See supra* note 193.

[260] *See supra* note 193.

[261] *See* Newman, *supra* note 257, at 860–63.

[262] Complaint at 26, Ass'n for Molecular Pathology v. U.S. Patent & Trademark Office, 702 F. Supp. 2d 181 (S.D.N.Y. 2010) (No. 1:09-cv-04515), 2009 WL 1343027, at *12–13.

[263] *See* Turna Ray, *With New BRCA Testing Options*, *Patient Groups Advise Women to Speak to Genetics Expert*, *Share Reports*, GENOMEWEB (June 19, 2013), https://www.genomeweb.com/clinical-genomics/new-brca-testing-options-patient-groups-advise-women-speak-genetics-expert-share [https://perma.cc/CM9X-5RL8] ("Immediately after the court released its decision, a number of labs announced that they would begin testing for BRCA alterations, including GeneDx, Pathway Genomics, Ambry Genetics, and Gene by Gene.").

[264] *See supra* notes 198–200 and accompanying text.

[265] *See Company Info*, FACEBOOK, http://newsroom.fb.com/company-info/ [https://perma.cc/5A8V-H2NB] (stating the amount of active Facebook users have grown to 1.71 billion despite any

Some users may not be aware that they, or more precisely their data, are the product.[266] Even consumers that appreciate their information is being sold in exchange for access to the service may undervalue that data.[267] Switching costs are presumptively lower here than in the genetic testing area, given the likelihood of overlapping patents and more limited coverage, allowing for more competition.[268] However, once users have contributed large amounts of information to a data-generating patent holder, such as Google or Facebook, they are likely to be reluctant to switch to providers who can only provide inferior search results or social networking.[269]

In this regard, although Google and Facebook may appear to be "free" to consumers, if that were the case, these companies would not earn any profit.[270] Rather, the users of Google and Facebook effectively pay these

concerns with the privacy policy as of June 2016); Vindu Goel, *Facebook Tinkers with Users' Emotions in News Feed Experiment, Stirring Outcry*, N.Y. TIMES (June 29, 2014), http://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html [https://perma.cc/H3NX-UV46] (explaining that in 2014, Facebook raised concerns when it was revealed that the social media network was manipulating their users' news feed for a study); Joshua Barrie, *Nobody Is Using Google+*, BUS. INSIDER (Jan. 20, 2015, 9:38 AM), http://www.businessinsider.com/google-active-users-2015-1 [https://perma.cc/7T8E-XEP7] (stating that approximately 9% of the 2.2 billion Google+ profiles are actively being used); Katy Barnato, *Google's Changes to Gmail Provoke More Privacy Fears*, CNBC (Jan. 10, 2014, 8:42 AM), http://www.cnbc.com/id/101326350 [https://perma.cc/U3SC-6BTF] (discussing how Google+'s privacy policy raised concerns in 2014 when it changed its policy so that anyone could look up any Gmail account without users giving their e-mail address out).

[266] *See* Steve Kovach, *Tim Cook Ripped Apart Google's Business Model in Two Paragraphs*, BUS. INSIDER AUSTL. (Sept. 18, 2014, 11:43 AM), http://www.businessinsider.com.au/tim-cook-privacy-letter-2014-9 [https://perma.cc/6ACH-HBTL]. In an open letter, Apple CEO Tim Cook explained that internet services, like Google, are using user's "personal data to market products to you." *Id.* Therefore, Cook explains that users of internet services are not the customer; rather, they are the product. *See id.*

[267] *See* Newman, *supra* note 257, at 860–63.

[268] *See* Motion for Preliminary Injunctive Relief and Memorandum in Support at 43, *In re* BRCA1– and BRCA2–Based Hereditary Cancer Test Patent Litig., 3 F. Supp. 3d 1213 (D. Utah 2014) (No. 2:13-cv-00640-RJS), 2013 WL 7862854, at *23 (stating that switching costs for genetic testing are very high because insurance companies will usually not "reimburse for a second, repetitive test"); Catherine Tucker & Alexander Marthews, *Social Networks, Advertising, and Antitrust*, 19 GEO. MASON L. REV. 1211, 1223 (2012) (stating that the switching costs for social networks like Facebook are fairly low).

[269] *See supra* note 136 and accompanying text; Shawn Baldwin, *Why Facebook Shares Will Rise Just Like Amazon's Did*, FORBES (Oct. 4, 2012, 12:49 PM) http://www.forbes.com/sites/shawnbaldwin/2012/10/04/facebook-2-0/#73b85e672f21 [https://perma.cc/448R-TJPN]; Monu Bedi, *Social Networks, Government Surveillance, and the Fourth Amendment Mosaic Theory*, 94 B.U. L. REV. 1809, 1880 (2014); William McGeveran, *The Law of Friction*, 2013 U. CHI. LEGAL F. 15, 58.

[270] Alejandro Crawford & Lisa Chau, *Why Google's Business Model Works,* U.S. NEWS (June 25, 2013, 10:35 AM) http://www.usnews.com/opinion/blogs/economic-intelligence/2013/06/25/why-googles-business-model-works [https://perma.cc/SY88-RWB5] ("Google makes the vast bulk of its revenues from AdWords."); Robert Hof, *Facebook's New Advertising Model: You*, FORBES (Nov. 16, 2011, 6:00 PM), http://www.forbes.com/sites/roberthof/2011/11/16/facebooks-new-advertising-model-

companies when they purchase products and services from advertisers, who in return remit a portion of that revenue back to Google and Facebook.[271] In a competitive market, these payments for products and services would presumably create no deadweight losses. However, because Google and Facebook enjoy market dominance, at least for some products and services, they arguably distort those markets to effectively eliminate competitive markets, engendering deadweight losses.[272] Additionally, deadweight losses arise from the opportunity costs consumers incur when high switching costs prevent them from adopting superior technologies.[273] To the extent data-generating patents contribute to high switching costs and market dominance, they concomitantly contribute to consumer deadweight losses.[274]

III. IDENTIFYING AND ADDRESSING PROBLEMATIC DATA-GENERATING PATENTS

Patent holders have likely captured data from their inventions—even in different product markets—since the beginning of the patent system in the fifteenth century. However, until recently, the collection and analysis of this data was generally quite costly and time-consuming, tempering social and economic concerns about leveraging patent protection to generate information protected by trade secret law. Now, in the era of big data, these limitations have been largely removed, and the potential for overreaching needs to be seriously reevaluated.[275] In the discussion below, we offer ways to distinguish problematic from unproblematic data-generating patents, and evaluate proposals to mitigate these potential concerns.

A. Discerning Problematic from Unproblematic Data-Generating Patents

As we explained earlier, not all data-generating patents necessarily result in net costs. In this Section, we propose criteria that can be used to identify potentially problematic data-generating patents. In our evaluation,

you/#572ed61852aa [https://perma.cc/J923-5EDM] (describing Facebook's use of advertising to generate revenue).

271 *See supra* note 270.

272 *See supra* Section II.C.2.

273 *See supra* Section II.C.2.

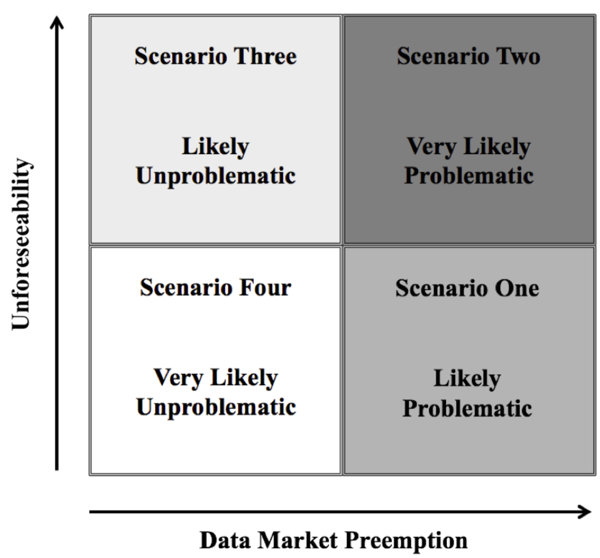
274 *See supra* Section II.C.2.

275 *See* Richards, *supra* note 159, at 1939 ("Big Data is notable not just because of the amount of personal information that can be processed, but because of the ways data in one area can be linked to other areas and analyzed to produce new inferences and findings."); *cf.* Simon, *supra* note 159 (explaining how the rise of "big data" techniques counsel in favor of modifying patent law's obviousness doctrine).

we focus on two main factors: expansion into unforeseeable markets and the strength of preempting potential competition in these markets. By unforeseeable markets, we refer to whether the patent affords the ability to collect data in an area that is not directly related to the market covered by the patented invention. If the invention allows aggregation of data in unforeseeable markets, the likelihood that it is problematic increases. For the preemption factor, we propose assessing the magnitude of the effect on competition in the market regarding the data, as opposed to the preemptive effect of the underlying invention itself. The greater the preemptive effect on marketplace competition related to the data, the more likely the invention is problematic. In many situations, these two factors will rise and fall together. However, as we show, many inventions may only implicate one of the factors. Moreover, although there are other factors that may be useful in such an analysis, such as the general strength of the patent holder in the marketplace, the commercial success of the invention, the size of the market, the type of technology involved, and whether the protected data implicates issues of major public concern, such as health, security, or privacy, we believe these two factors tend to do much of the work in determining whether a given data-generating patent is likely to be problematic.

We discuss four scenarios in ascertaining whether a data-generating invention may cause concern. Each scenario turns on the level of unforeseeability and preemption present in a given data-generating patent (or set of patents). See Figure 1.

FIGURE 1: DATA-GENERATING PATENT SCENARIOS



1. Scenario One (High Preemption, Low Unforeseeability).—In the first scenario, the invention at issue allows collection of data in markets that are closely related to the invention, but the preemptive effect of the patent may be so great that the costs of allowing the patent holder to maintain the data generated by the invention secret may be unduly high.

For example, take the Myriad patent related to screening for susceptibility to breast and ovarian cancer.[276] Myriad's use of the data it collects to improve on its invention, including reducing the number of variants of unknown significance, is arguably directly related to the market of the underlying invention.[277] Even using the data to engage in additional cancer research and discovery, such as other types of breast cancer, would be closely related to the area of the patented invention, and thus foreseeable.[278] Google's use of data obtained through the patented PageRank technology to provide personalized search results, for example, is similar to the Myriad case in that it is arguably foreseeable.[279]

The main reason that these types of inventions may be problematic, however, is their highly preemptive nature. Such preemption may or may not turn heavily on the presence of patent protection. When broad preemptive scope is directly related to relevant data-generating patents, absent technological strides to design around the scope of the patents, the effect on competition in the market regarding the data is essentially preclusive during the twenty-year exclusivity period.[280] Although aggregating data may lead to increased efficiencies, it may also result in more limited data analysis and improvements resulting from the data by the patentee or third parties.[281] Additionally, when the ability to generate the data is exclusively in the hands of one provider, the ability to assess the quality of any claims by the data aggregator is limited.[282] As discussed earlier, after the patent expires or is invalidated, later entrants play a costly and extensive game of catch-up.[283]

In terms of the additional factors we mention above, the patent holder of a data-generating patent with a highly preemptive effect will usually

---

276 *See supra* Section II.A.1.

277 *See supra* Section II.A.1.

278 *See generally* Eleonore Pauwels, Opinion, *Our Genes, Their Secrets*, N.Y. TIMES (June 18, 2013), http://www.nytimes.com/2013/06/19/opinion/our-genes-their-secrets.html?_r=0 [https://perma.cc/VZ4T-XGZV] (describing how Myriad's database might be used for related cancer research).

279 *See supra* Section II.A.2.

280 *See supra* Section II.B.2.

281 *See supra* Section II.B.3.

282 *See, e.g.*, Simon, *supra* note 99, at 1317 (describing how patents can prevent third-party assessment of an invention).

283 *See supra* Section II.C.2.

have unparalleled strength in the marketplace.[284] The commercial success of the invention and the size of the market will depend on the value that the data-generating patent provides to interested consumers.[285] In the case of genetic testing, the market will depend not only on the usefulness of the patent to consumers, but also on the likelihood of insurance coverage.[286] With regard to the technology involved, the data generated in the genetic testing space will have an impact on privacy and public health, which are serious public concerns to consider in determining if the data-generating patent is problematic.[287] In general, determining whether a data-generating patent in this scenario is problematic will often require resorting to these kinds of other factors.[288]

2. Scenario Two (High Preemption, High Unforeseeability).—In the second scenario, the invention at issue allows the patent holder to gather data in an unforeseeable market and the preemptive effect of the patent is substantial. Altering a commonly referenced bioethics case study for purposes of this discussion, imagine the possibilities for generating data from a patent related to genetic testing for susceptibility to diabetes.[289] If the patent holder used the gathered data to determine that a certain ethnic group had a propensity for developing another disease, such as muscular dystrophy or schizophrenia,[290] the data collected would reach unforeseeable markets, unlike the Myriad situation.[291] If the patent were of strong preemptive effect, this would afford the patent holder market power in independent markets during and even after the expiration or invalidation of the patent.[292] This unforeseeable market power is, by definition, not

---

284 *See supra* note 95 and accompanying text; *supra* Section II.A.1.

285 *See supra* note 259 and accompanying text.

286 *See* Motion for Preliminary Injunctive Relief and Memorandum in Support at 43, *In re* BRCA1– and BRCA2–Based Hereditary Cancer Test Patent Litig., 3 F. Supp. 3d 1213 (D. Utah 2014) (No. 2:13-cv-00640-RJS), 2013 WL 7862854, at *23 (discussing the effect of insurance coverage on switching costs for genetic testing).

287 *See* Barbara J. Evans, *Much Ado About Data Ownership*, 25 HARV. J.L. & TECH. 69 (2011) (discussing data access and privacy concerns related to public uses of health data).

288 *See supra* notes 284–87 and accompanying text.

289 For purposes of discussing this altered case study, we have also assumed that the researchers obtained informed consent from participants. *See* Michelle M. Mello & Leslie E. Wolf, *The Havasupai Indian Tribe Case—Lessons for Research Involving Stored Biologic Samples*, 363 NEW ENG. J. MED. 204 (2010) (discussing claims that Arizona State University "researchers improperly used tribe members' blood samples in genetic research"); Amy Harmon, *Indian Tribe Wins Fight to Limit Research of Its DNA*, N.Y. TIMES (Apr. 21, 2010), http://www.nytimes.com/2010/04/22/us/22dna.html?pagewanted=all&_r=1 [https://perma.cc/PG2P-SFLZ] (describing allegations that researchers used Havasupai blood to study things other than diabetes, such as mental illness).

290 *See supra* note 289.

291 *See supra* Section II.A.1.

292 *See supra* note 84.

contemplated in the scheme of patent rewards and corresponding social costs erected by the patent system.[293] Given the possibility of over-rewarding the patent holder in the face of heightened deadweight consumer losses, as well as potential downstream barriers to innovation, patents falling into this category present the largest potential threat to social welfare.[294]

Similarly, Google's use of generated data to engage in targeted advertising, for example, may provide it market power wholly unrelated to search technology.[295] By leveraging its PageRank technology to obtain a wealth of data about its users, Google is able to reach into unforeseeable secondary markets to offer its business partners a superior way to engage in targeted advertising of its users.[296] If the PageRank patent is highly preemptive in gathering the user data, considering Google's dominance in the search industry, this use of data-generating patents presents potentially serious policy concerns.[297]

With regard to the other potentially relevant factors, the patent holder of a data-generating patent with a highly preemptive effect that allows for expansion into secondary markets will usually have dominance in the primary (and perhaps secondary) marketplace.[298] Because of their preemptive and expansive nature, these patents will often be commercially successful and affect participants in more than one market.[299] As to the technology involved, the data generated in both genetic testing and the search industry may raise substantial privacy concerns.[300]

In sum, the ability to keep the aggregated data as a trade secret is likely to prove more problematic than scenario one, because it presents the possibility of foreclosing competition not only in the primary market of the invention, but also in secondary markets. Expanding the reach of the patent into unforeseeable markets seems particularly problematic because it

---

293 *See supra* Section II.B.

294 *See supra* Section II.C.

295 *See supra* note 145.

296 *Supra* note 145. One might wonder why targeted advertising is not a "foreseeable" market extension of internet search technology. As noted earlier, by "unforeseeable" we use the narrow legal connotation of whether the secondary market is directly related to the primary market. *See supra* note 23 and accompanying text. In this sense, targeted advertising would not be a foreseeable extension of internet search.

297 *See supra* Section II.A.2.

298 *See supra* note 95 and accompanying text; *supra* Section II.A.1.

299 *See supra* note 95 and accompanying text; *supra* Section II.A.1.

300 *See supra* note 194; Omer Tene, *What Google Knows: Privacy and Internet Search Engines,* 2008 UTAH L. REV. 1433, 1435–38.

allows the data-generating patent holder not merely to improve upon the patented invention, but broadly to foreclose competition.

3. Scenario Three (Low Preemption, High Unforeseeability).—For the third scenario, the invention may only have a weak preemptive effect, but the invention allows the patent holder to gather data in an unforeseeable market. For example, in the social networking space, Facebook has acquired patents related to providing a news feed, though they are arguably quite narrow.[301] These patented inventions have enabled Facebook to obtain information related to its users' browsing patterns, for example.[302]

For one week in 2012, using its patented techniques, Facebook directed the news feeds of almost 700,000 of its users to alter the number of negative posts they viewed, using the data it collected to evaluate whether emotions were contagious on social media.[303] Facebook never obtained explicit consent for the experiment, relying instead on its users' prior agreement to its Terms of Service, causing an uproar among many commentators and users.[304]

In this sense, the Facebook newsfeed patent allows the gathering of data in unforeseeable markets, such as studying the effects of negative inputs on emotion, although it is apparently not particularly preemptive given its seemingly narrow scope.[305] In simpler terms, there are other viable ways to develop the aggregated data, such as through other social networking and news media sites, as well as university or industry research.[306]

The patent holder of a data-generating patent that allows it to reach into secondary markets, but with low preemptive effect, does not tend to

---

301 *See, e.g.*, U.S. Patent Nos. 7,669,123 (filed Aug. 11, 2006) (issued Feb. 23, 2010) (describing Facebook's newsfeed technology); 7,827,208 (filed Aug. 11, 2006) (issued Nov. 2, 2010) (disclosing creation of personalized news feeds).

302 *Data Policy*, FACEBOOK, https://www.facebook.com/policy.php [https://perma.cc/GX8Q-CWBN] (last modified Sept. 29, 2016).

303 Adam D.I. Kramer et al., *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks,* 111 PNAS 8788 (2014), http://www.pnas.org/content/111/24/8788.full [https://perma.cc/CJA7-LJDL]; Goel, *supra* note 265.

304 Goel, *supra* note 265.

305 *See supra* note 146 and accompanying text.

306 *See, e.g.*, Alexander H. Jordan et al., *Misery Has More Company Than People Think: Underestimating the Prevalence of Others' Negative Emotions*, 37 PERSONALITY & SOC. PSYCHOL. BULL. 120–35 (2011), http://psp.sagepub.com/content/37/1/120.abstract [https://perma.cc/8UY8-SDED]; Andrew Perrin, *Social Media Usage: 2005-2015*, PEW RESEARCH CTR. (Oct. 8, 2015), http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [https://perma.cc/CR6V-K5QL].

exert dominance in primary or secondary markets.[307] Although these inventions will often be commercially successful, they are unlikely to raise significant concerns because the leverage they provide is minimal.[308] If the preemptive effect of such a patent is low enough so that it provides effectively no leverage to the patent holder in collecting data maintained as a trade secret, then it would be no more a concern than collecting data in the absence of a patent.[309] Indeed, the uproar over Facebook's manipulation of user sentiment arguably had little to do with its newsfeed patent but rather with core privacy issues.[310] However, as the preemptive capabilities rise, the ability to collect data in unforeseen markets presents greater concerns, approaching the likely problematic effects of patents falling under scenario two (high preemption and high unforeseeability).

4. Scenario Four (Low Preemption, Low Unforeseeability).—In this final scenario, an invention with a weak preemptive effect that allows for gathering data in a foreseeable market is unlikely to be problematic. For instance, a narrow patent covering a very specific method of conducting an online survey is unlikely to be preemptive, as there are many alternative methods of conducting a survey online.[311] Similarly, because surveys are designed to gather a wide variety of data, it is unlikely that any use for which data is gathered would be unforeseeable.[312] The use of the data in this way is closely related to the area covered by the patent, and because it is not particularly preemptive, it is unlikely to impose social costs in excess of its benefits (assuming it was properly granted).

B. Proposals to Address Problematic Data-Generating Patents

We have discussed concerns related to problematic data-generating patents; in this Section, we offer some ways to mitigate their detrimental effects. In particular, we suggest and assess a variety of means to address potentially detrimental effects of data-generating patents on innovation and

---

[307] *See* Mark A. Lemley, Michael Risch, Ted Sichelman & R. Polk Wagner, *Life After* Bilski, 63 STAN. L. REV. 1315, 1344–45 (2011) (describing how broad patent claims can preempt all uses of the claimed invention).

[308] *See id.*

[309] *See id.*

[310] *See* Kramer et al., *supra* note 303; Goel, *supra* note 265.

[311] *See, e.g.*, *Collecting Survey Data*: *Internet Surveys,* PEW RESEARCH CTR., http://www.pewresearch.org/methodology/u-s-survey-research/collecting-survey-data/ [https://perma.cc/L3DG-S3BK].

[312] Even though foreseeable, these data markets would not concern the invention itself. Thus, to the extent the patents were preemptive—specifically, foreclosing alternative methods of gathering such data—it could be quite problematic. *See supra* note 23 and accompanying text.

disclosure.[313] We evaluate the effects of limiting patentability, mandating narrow disclosure or sharing of data, expanding defenses to infringement, and restricting available remedies. Although none of the many potential remedies we discuss are without downsides, we tend to favor ex post solutions—either through private or public actions in the courts, or a regulatory arena—that can be tailored to the specific circumstances of a problematic data-generating patent.[314]

1. Mitigating Innovation-Related Concerns.—As we discussed earlier, prohibitive switching costs may prevent consumers from choosing new market entrants after the patent expires or is invalidated, extending deadweight losses.[315] Furthermore, concentrating data in one provider may hinder improvements, particularly where the patent holder has diminished incentives to further innovate, such as when investment in the existing infrastructure yields satisfactory financial returns.[316]

We suggest and consider some alternatives to mitigate these harmful effects when they occur. As an initial matter, not all data-generating patents are cause for concern.[317] Rather, it is only a subset of those patents—primarily those that are preemptive in unforeseeable markets (and sometimes those that are highly preemptive in foreseeable markets)—that may impose substantial social costs.[318] Identifying these problematic patents ex ante—for instance, at an agency such as the USPTO—will likely be costly, difficult, time consuming, and error-ridden.[319]

Substantially shortening the patent term or wholly eliminating patent protection on data-generating patents would ease the burden at the USPTO,

---

313 *See supra* Part II. In what follows, we assume that the underlying patents were at least arguably valid when granted; of course, the acquisition of market power in the presence of invalid patents generally results in deadweight losses without sufficient corresponding gains. *See* Anup Malani & Jonathan S. Masur, *Raising the Stakes in Patent Cases*, 101 GEO. L.J. 637, 639 (2013) ("[I]nvalid patents have no upside: they do not encourage innovation, and they impose deadweight losses on welfare."). Nonetheless, invalid data-generating patents can exacerbate the problems we discussed previously by allowing the patent holder to retain data as a trade secret, even after invalidation. *See supra* Section II.B.

314 *See* Ian Ayres & Eric Talley, *Solomonic Bargaining: Dividing a Legal Entitlement to Facilitate Coasean Trade*, 104 YALE L.J. 1027, 1092 (1995) (discussing the role of the courts in the "*ex post* tailoring" of remedies).

315 *See supra* notes 211–13 and accompanying text.

316 *See* Merges & Nelson, *supra* note 63, at 843, 907–09; Saunders, *supra* note 246, at 423; Simon, *supra* note 99, at 1317.

317 *See supra* Section III.A.

318 *See supra* Section III.A.

319 *See* Lemley, Risch, Sichelman & Wagner, *supra* note 215, at 1326 (noting how in the context of patentable subject matter determinations "gatekeeping theories are necessarily bright-line rules: they will both exclude some patents that should be granted and fail to exclude others that should not").

but is unlikely to be optimal.[320] A restricted exclusivity period would cabin the patent holder's ability to gather data, but in addition to classification problems, it also might have the unintended consequence of nudging providers to maintain their data-generating inventions as trade secrets from the start, rather than seeking patent protection.[321] To be certain, forcing data-generating inventions into trade secrecy levels the playing field so that these inventors do not enjoy "super-patent" protection that allows them to use market power conferred by the patent to achieve dominance in the market for data protected by trade secret law.[322] In other words, in a pure trade secrecy regime, competitors can all use the offensive and defensive aspects of trade secret law—including reverse engineering and independent discovery—which arguably increases innovation incentives and drives down prices.[323]

On the other hand, trade secrecy not only would reduce disclosure but also could cause providers to take extra precaution in preventing leakage of proprietary data-generating methods, such as through increased vertical integration, increased usage of nondisclosure and noncompete agreements, and segmenting of employees.[324] These negative effects could potentially reduce innovation more than the social costs otherwise imposed by data-generating patents. Coupled with the difficulty of classifying harmful data-generating patents ex ante, shortening patent term does not appear to be a viable solution.

Indeed, for some classes of data-generating inventions, as we discussed earlier, the courts have already chosen a similar, but even more drastic solution: eliminating patent protection altogether. These categories include most medical diagnostic inventions as well as most inventions embodying algorithms, which likely includes Google's PageRank methods.[325] Eliminating patent eligibility for data-generating inventions, even harmful ones, would simply exacerbate the problems of shortening

[320] *Cf.* Brian J. Love, *An Empirical Study of Patent Litigation Timing: Could a Patent Term Reduction Decimate Trolls Without Harming Innovators?*, 161 U. PA. L. REV. 1309, 1313 (2013) (suggesting that Congress could "shorten the patent term by three years or even longer" without harming innovation).

[321] *See generally* Schwartz, *supra* note 41, at 637, 648–52 ("[T]he obvious preference for an IP holder is for a long term . . . .").

[322] *See supra* note 189 and accompanying text.

[323] *See supra* Section II.B.

[324] *See* Jonathan M. Barnett & Ted Sichelman, *Revisiting Labor Mobility in Innovation Markets* (working paper, May 26, 2016), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2758854 (describing the role of noncompete and non-solicitation agreements and alternatives such as vertical integration and segmenting employees).

[325] *See supra* Section II.B.3.

patent term. Of course, there may be other sound reasons for eliminating these classes of patentable subject matter,[326] but with respect to reducing the costs of harmful data-generating patents—and there certainly are classes of these inventions that remain patentable[327]—we do not believe such an approach is sensible.

A second solution would extend the patent misuse defense to bar enforcement of data-generating patents that improperly expand the scope of the underlying legal rights. This approach would not be as harsh as eliminating patentability entirely for data-generating inventions, but could yield similar effects.[328] Specifically, to the extent that would-be patentees for problematic data-generating inventions could predict that their patents would be precluded from enforcement, they would opt for trade secrecy, presenting similar concerns to those discussed earlier.[329] So while expanding misuse is preferable to all-encompassing solutions, it is far from ideal.[330]

A third option would be the expansion of the independent invention exception to patent infringement for data-generating patents.[331] This approach would allow competitors to generate databases similar to the data-generating patent holder, as long as competitors did not reverse engineer or copy data from the patent holder.[332] Congress has successfully enacted narrow exceptions to allow for independent invention under the America Invents Act.[333] However, the independent invention must be in commercial use as a process or as a product in a manufacturing process at least one year prior to the filing of the patent at issue.[334] These limitations

[326] *See supra* note 49 and accompanying text.

[327] *See supra* Section II.A.3.

[328] *See supra* note 227 and accompanying text.

[329] *See supra* notes 320–25 and accompanying text.

[330] Similarly, expanding antitrust causes of action to allow for affirmative counterclaims to enforcement of problematic data-generating patents would not be an optimal solution in our view. *See* USM Corp. v. SPS Techs., Inc., 694 F.2d 505, 512 (7th Cir. 1982) (Posner, J.) ("If misuse claims are not tested by conventional antitrust principles, by what principles shall they be tested? Our law is not rich in alternative concepts of monopolistic abuse; and it is rather late in the day to try to develop one without in the process subjecting the rights of patent holders to debilitating uncertainty."); Barbara Evans, *Economic Regulation of Next-Generation Sequencing,* J.L. MED. & ETHICS, Fall 2014, https://www.aslme.org/media/downloadable/files/links/0/5/05.SUPP_Evans.pdf [https://perma.cc/RD5D-4FYW] (describing the application of antitrust law to the "data hoarding" problems associated with holders of genetic testing patents).

[331] *See, e.g.*, Samson Vermont, *Independent Invention as a Defense to Patent Infringement*, 105 MICH. L. REV. 475 (2006) (proposing a general independent invention defense to patent infringement).

[332] *See id.*

[333] 35 U.S.C. § 273 (2012).

[334] *See id.*

could be eliminated so as to allow a general independent invention defense before or even after the filing of the relevant patent.[335] A benefit of an independent invention defense is that it is applied ex post in litigation, so courts could tailor the remedy to target socially harmful data-generating patents. For instance, the independent invention provision could be amended only to apply to "reasonable" uses of the invention, and courts could adopt a factor-based test—along the lines of that described earlier—to determine such uses.[336]

Indeed, one could even expand the defense to allow reverse engineering of the patented invention so long as it was used to generate data that otherwise could not be collected without use of the invention.[337] For example, consider a patent with substantial preemptive effects, such as one related to genetic markers, where the exclusivity afforded by the data-generating patent might foreclose competition for gathering related data.[338] While a reverse engineering exception would allow for competition in the market of related genetic data, permitting even a limited exception would entail large monitoring costs to ensure that competing data gatherers were not using reverse engineering as a pretext for engaging in infringing behavior.[339] Monitoring the use of patented inventions in this way could be challenging, though Congress has adopted similar exceptions in the past, such as the use of certain patented inventions for reasons related to the submission of information to the Food and Drug Administration (FDA).[340]

Fourth, for data-generating patents that restrict basic research, such as Myriad's withholding of genotypic and phenotypic information from public databases,[341] Congress or the courts could adopt an experimental use exception to patent infringement for data-generating patents.[342] Again, the exemption related to FDA approval as well as the exception for medical and surgical procedures performed by physicians are potential models for a

[335] *See* Vermont, *supra* note 331, at 484–92.

[336] *See supra* Section III.A.

[337] *See supra* Section III.A.

[338] *See supra* Section III.A.

[339] *See* Pamela Samuelson & Suzanne Scotchmer, *The Law and Economics of Reverse Engineering*, 111 YALE L.J. 1575, 1650–51 (2002).

[340] *Cf.* 35 U.S.C. § 271(e)(1) (2012).

[341] *See supra* note 262 and accompanying text.

[342] *See, e.g.*, Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017, 1074–78 (1989) (discussing "the proper scope of an experimental use exemption"); Maureen A. O'Rourke, *Toward a Doctrine of Fair Use in Patent Law*, 100 COLUM. L. REV. 1177, 1203–11 (2000) (proposing a fair use exception for patent law).

research-based exemption.[343] Similar experimental use exceptions are also recognized internationally.[344] While Congress might not adopt such an exception, the courts could plausibly do so, though they generally have been hesitant to broaden the scope of infringement exceptions for research.[345] Like an independent invention defense, a research exemption could be tailored by the courts to apply only to harmful data-generating patents.[346]

One downside of the independent invention and experimental use defenses is that they are complete bars to enforcement, at least in a particular lawsuit. If the data-generating patentee is effectively unable to enforce its patent against most infringers, such defenses effectively function as a denial of patentability altogether. In this situation, the result is similar to that described earlier—increased incentives for the inventor to retain its invention as a trade secret, with the attendant costs of doing so.

Another solution that can be applied in litigation—and, thus, tailored to harmful data-generating inventions—is placing limits on patent remedies.[347] Courts, for instance, could limit injunctive relief for infringing these types of patents, given the difficulty of reverse engineering and independently creating the data generated from them during the exclusivity period.[348] Considering that these safeguards of trade secret law are virtually nonexistent for these types of inventions, courts could carefully adopt a

[343] § 271(e)(1) (creating an infringement exception for activities reasonably related to FDA submission); *id.* § 287(c) (prohibiting patent enforcement of medical and surgical procedures against physicians); *id.* § 273 (expanding the prior use defense under previous law). *See generally* Peter S. Menell, *Tailoring a Public Policy Exception to Trade Secret Protection*, 105 CALIF. L. REV. (forthcoming), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2686565 [https://perma.cc/6AYH-8LJL] (proposing an exception to trade secret protection for the reporting of illegal activity to trusted intermediaries).

[344] *See, e.g.*, John F. Duffy, *Harmony and Diversity in Global Patent Law*, 17 BERKELEY TECH. L.J. 685, 718–19 (2002) ("The United Kingdom, Germany, Japan, Korea and many others expressly recognize an experimental use exception in their statutory law.").

[345] *See, e.g.*, Madey v. Duke Univ., 307 F.3d 1351, 1362 (Fed. Cir. 2002) (refusing to expand the research exemption to nonprofit activity).

[346] *See supra* note 333 and accompanying text.

[347] *See* Ted Sichelman, *Purging Patent Law of "Private Law" Remedies*, 92 TEX. L. REV. 517 (2014) (discussing reforms in the area of patent remedies); John M. Golden, *Principles for Patent Remedies*, 88 TEX. L. REV. 505, 556–57 (2010) (discussing various approaches to patent remedies).

[348] *See supra* Section III.A. Of course, courts could also limit damages, but determining the exact diminution would be a difficult task. *See* Sichelman, *supra* note 347, at 553. Relatedly, for medical diagnostic tests, governments could refuse to reimburse data-generating patent holders—at least where the patent has expired or has been invalidated—for performing medical tests absent disclosure of the underlying data, for instance, to a public database. *See* Cook-Deegan et al., *supra* note 37, at 587. Such a solution is again fairly draconian, and would likely shift the protection of data-generating inventions into trade secrecy. *See supra* notes 320–25 and accompanying text. As such, we are generally averse to altering medical payment schemes. *See supra* notes 320–25 and accompanying text.

presumption against granting an injunction, particularly in assessing whether the balance of hardships tilts in the infringer's favor and the potential harm to the public interest.[349] Following the Supreme Court's decision in *eBay Inc. v. MercExchange, L.L.C.*, courts have frequently denied injunctions to non-practicing entities (NPEs) and sometimes to practicing entities, particularly in the field of information technologies, in order to prevent "holdup" and other problems flowing from injunctive relief.[350] If courts can properly take into account the costs caused by data-generating patents, yet still award monetary relief sufficient for inventors to seek patent protection, this solution could be ideal.[351] On the other hand, systematic errors in the award of monetary relief could result, like other solutions described earlier, in de facto patent ineligibility for many data-generating inventions. This result would be especially concerning in an area where courts have already significantly contracted patent eligibility.

2. Addressing Disclosure-Related Concerns.—As we have explained, as a result of patent exclusivity, data-generating inventions capture information that will not fall into the public domain after the patent has expired or is invalidated. Some of the same potential solutions to data-generating patents' innovation-related concerns would mitigate these disclosure problems. For instance, the ability to reverse engineer a data-generating patent would spur the production of otherwise secret information by competitors. Here, we examine solutions that more directly address disclosure issues.

One possibility is to force the holder of a harmful data-generating patent to publicly disclose the data otherwise protected by trade secret law.[352] As an initial matter, it would be difficult to enforce such a mandate,

---

[349] *See* eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388, 391–94 (2006).

[350] *See* Christopher B. Seaman, *Permanent Injunctions in Patent Litigation After* eBay*: An Empirical Study*, 101 IOWA L. REV. 1949, 2002–06 (2016).

[351] *See* Sichelman, *supra* note 347, at 520–29 (suggesting eliminating injunctive relief for a broad class of patents and parties); Barbara J. Evans, *Mining the Human Genome After* Association for Molecular Pathology v. Myriad Genetics, 16 GENETICS IN MED. 504, 506 (2014) (discussing liability rule regimes in the context of biomedical data); J.H. Reichman & Paul F. Uhlir, *A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment*, 66 LAW & CONTEMP. PROBS. 315, 350 (2003) ("[T]he generally weak liability rules governing confidential information can produce too much protection for data held in actual secrecy because there is, in practice, virtually no functional equivalent of reverse engineering.").

[352] Such an approach is related to Gideon Parchomovsky and Michael Mattioli's proposal of the "semi-patent." Gideon Parchomovsky & Michael Mattioli, *Partial Patents*, 111 COLUM. L. REV. 207, 208 (2011). Specifically, the grant of the semi-patent would be conditioned on the applicant's agreement to publish its research data, both positive and negative, as it relates to the patented invention. *See id*. Our proposal, however, is not that the patent holder be required to publish research data about

because the patent holder could disclose a portion of its proprietary data, and it would be very costly for regulators to determine whether additional data was not disclosed.[353] Moreover, without protecting that information as a trade secret, the patent holder would have less incentive to generate the information, potentially resulting in even less information for the public to use than with data-generating patents.[354] Plus, the development of information that is available may come at a higher cost.[355]

A less draconian way to encourage information sharing would be to mandate donation of data produced as a result of a data-generating patent to a regulatory agency, such as to the FDA or NIH, which would not disclose it for a certain period of time.[356] A limited period of data exclusivity might provide sufficient innovation incentives in this space, much as it has for biologics.[357] Generally, confidential data submitted during the FDA approval process remains behind locked doors.[358] Again, some mechanism to detect data-generating patents, or even better, problematic data-generating patents would need to be adopted ex ante, and it may simply be too difficult to do so for most technology areas.[359] However, patents in some technological fields—such as genetic diagnostic tests and internet search processes—could include such a high percentage of data-generating patents that the costs of classification (and evasion of such by savvy patent applicants) may be dwarfed by the benefits of disclosure.[360]

---

the invention; rather, we propose limited disclosure of the independent data generated by use of the invention. *See infra* notes 353–78 and accompanying text.

[353] *See, e.g.*, Michael Mattioli, *Data Pools*, 32 BERKELEY TECH. L.J. (forthcoming 2017), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2671939 [https://perma.cc/8KRY-JTHU].

[354] *See* Lichtman, *supra* note 71, at 230.

[355] *See id*.

[356] *See, e.g.*, National Institutes of Health, Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. 51,345–54 (2014), https://www.federalregister.gov/articles/2014/08/28/2014-20385/final-nih-genomic-data-sharing-policy [https://perma.cc/FRM6-5JWL]; Barbara J. Evans et al., *The FDA and Genomic Tests—Getting Regulation Right*, 372 NEW ENG. J. MED. 2258, 2258, 2262 (2015) (concluding that although the FDA currently lacks "the correct set of statutory powers to make genomic technologies safe and effective . . . while still fostering innovation," implementing modest statutory reforms could "encourage public–private partnerships to develop and sustain data resources for the right regulation of genomic testing"); Price II, *supra* note 58.

[357] *See* Yaniv Heled, *Patents vs. Statutory Exclusivities in Biological Pharmaceuticals—Do We Really Need Both?*, 18 MICH. TELECOMM. & TECH. L. REV. 419, 449–61 (2012) (describing the implications of having statutory exclusivities and patent protection for biologics).

[358] 21 U.S.C. § 331(j) (2012) (prohibiting the disclosure of "any method or process which as a trade secret is entitled to protection").

[359] *See supra* notes 317–23 and accompanying text.

[360] *See* Lemley, Risch, Sichelman & Wagner, *supra* note 215, at 1327 ("[G]atekeeping rules don't have very clear lines; subject matter category delineation is notoriously elusive . . . .").

In the area of federally funded research, agencies could mandate such limited disclosure without intervention from Congress.[361] To foster disclosure, some have suggested amending the Bayh-Dole Act or increasing the authority of the NIH to ensure availability of information that is critical to public well-being.[362] Recently, the NIH has instituted requirements for sharing human genomic data generated by NIH-funded research, although the data remains sequestered for a substantial period of time after submission, after which point it is made available through a tiered distribution system.[363] Nonetheless, more stringent disclosure requirements would face strong industry opposition, monitoring and compliance challenges, and may too greatly affect innovation incentives,[364] although empirical evidence is lacking to draw a conclusion.[365]

Along these lines, public or public–private partnerships could incentivize the collection and analysis of data.[366] Considering the use of

---

361 *See* STEPHEN BREYER ET AL., ADMINISTRATIVE LAW & REGULATORY POLICY (5th ed. 2001).

362 *See* Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 LAW & CONTEMP. PROBS. 289, 294 (2003) (discussing administrative obstacles the NIH faces in intervening); Andrew W. Torrance, *Patents to the Rescue - Disasters and Patent Law*, 10 DEPAUL J. HEALTH CARE L. 309, 341 (2007) ("[T]o date no such petition has resulted a grant of march-in rights by the federal government."); Sara Boettiger & Alan B. Bennett, *Bayh-Dole: If We Knew Then What We Know Now*, 24 NATURE BIOTECHNOLOGY 320, 323 (2006) (noting that attempts to amend Bayh-Dole typically have not succeeded). Although the Bayh-Dole Act does not cover data per se, as we have explained, patents may effectively block full use of the data. *See supra* Section II.A; Rebecca S. Eisenberg & Arti K. Rai, *Harnessing and Sharing the Benefits of State-Sponsored Research: Intellectual Property Rights and Data Sharing in California's Stem Cell Initiative*, 21 BERKELEY TECH. L.J. 1187, 1189 (2006) ("[T]he Bayh-Dole Act . . . does not directly address the dissemination of unpatentable data . . . ."). So although federal agencies could mandate release of data that result from federally funded projects, full use of the data by third parties would in some cases require the Bayh-Dole Act to be amended to allow use of related patents. *See supra* Section II.A.

363 National Institutes of Health, Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. 51345-54 (2014), https://www.federalregister.gov/articles/2014/08/28/2014-20385/final-nih-genomic-data-sharing-policy [https://perma.cc/TS6L-5FT4].

364 *See* Saunders, *supra* note 246, at 439–40 (discussing concerns about reducing incentives in disclosure and the development of inventions).

365 Along these lines, but rarely used in the United States, compulsory licensing of the data could provide for broader information sharing. *See, e.g.*, Torrance, *supra* note 362, at 336–40. Again, however, such approaches could unduly reduce incentives to patent, channeling data-generating inventions into trade secrecy. *See supra* notes 320–25 and accompanying text. Additionally, requiring the disclosure of proprietary data may, under some circumstances, constitute a taking requiring just compensation. *See* Ruckelshaus v. Monsanto Co., 467 U.S. 986, 1003, 1005 (1984) (finding that Monsanto's "interest in its health, safety, and environmental data [is] cognizable as a trade secret property right" and discussing "several factors that should be taken into account when determining whether a governmental action" rises to the level of a taking).

366 *See, e.g.*, Evans, *supra* note 330 (evaluating a variety of public and private efforts with regard to data sharing); DEP'T OF HEALTH & HUMAN SERVS., HEALTH DATA INITIATIVE: STRATEGY AND EXECUTION PLAN 9 (2013), https://www.healthdata.gov/blog/health-data-initiative-strategy-execution-plan-released-and-ready-feedback [https://perma.cc/2W4U-JRGJ] (describing the Health Data

data about consumers as part of the public infrastructure would encourage investment in its development.[367] Congress could enact legislation or the courts could impose a remedy in trade secret law similar to the "essential facilities" doctrine in antitrust to ensure that markets and the public would have access to critical information.[368] Like forcing information disclosure, however, it would seem too difficult to apply such an approach in practice to all but a narrow class of inventions.[369] Another option would be for the government to encourage the voluntary adoption of such sharing arrangements.[370] These forms of data aggregation are rare in the United States, but they have enjoyed more success internationally; perhaps financial and other encouragement could increase their use domestically.[371] Nonetheless, given the strong financial incentives to avoid such voluntary arrangements, and difficulty in standardization, we do not view them as likely to solve a major share of problems created by harmful data-generating patents.[372]

A sui generis form of protection for data, such as that adopted by the European Union, would potentially encourage data-generating patent holders to disclose their information and charge for its use.[373] However, database protection in Europe is typically in the context of information that must be disclosed for its holder to monetize it—for instance, a database of

---

Consortium); Press Release, Ambry Genetics, Ambry Genetics Launches AmbryShare, the Largest Disease-Specific Public Database of Sequenced Human Genomes, Increasing the Understanding of Hereditary Breast and Ovarian Cancer 10 Fold (Mar. 7, 2016), https://share.ambrygen.com/news/detail/1/ambry-genetics-launches-ambryshare-the-largest-disease-specific-public-database-of-sequenced-human-genomes-increasing-the-understanding-of-hereditary-breast-and-ovarian-cancer-10-fold [https://perma.cc/G5F8-82DB] (releasing "anonymized, aggregated data of 10,000 human genomes" and committing to "potentially contributing data from almost 200,000 genomes annually based on projections").

367 *See* Price II, *supra* note 89, at 419; BRETT M. FRISCHMANN, INFRASTRUCTURE: THE SOCIAL VALUE OF SHARED RESOURCES 137 (2012).

368 *See* Evans, *supra* note 330, at 62 (noting that "the doctrine [is] technically alive, albeit eviscerated, in Supreme Court jurisprudence").

369 *See supra* notes 357–65 and accompanying text.

370 *See* Cook-Deegan, *supra* note 37, at 587–88.

371 *See supra* notes 356–67 (examining structures to encourage data sharing); Francis S. Collins et al., *The Human Genome Project: Lessons from Large-Scale Biology*, 300 SCIENCE 286–90 (2003) (discussing the race to sequence the human genome by the Human Genome Project, a public–private partnership, and Celera Genomics, a private firm).

372 *See, e.g.*, Mattioli, *supra* note 89, at 545 ("[D]ata is often recorded and published in a wide variety of formats."); Turna Ray, *In Tackling the VUS Challenge, Are Public Databases the Solution or a Liability for Labs*?, GENOMEWEB (Feb. 12, 2014), https://www.genomeweb.com/clinical-genomics/tackling-vus-challenge-are-public-databases-solution-or-liability-labs [https://perma.cc/U59E-LLQP] (discussing how some public variant genomic databases do not "adopt a consistent evidence-based standard for variant interpretation" and are not updated regularly).

373 *See* Mattioli, *supra* note 89, at 580.

names, addresses, and emails for sale to marketers.[374] Given that the information produced by data-generating patents can be profitably maintained as a trade secret, it is unlikely that sui generis protection would affect behavior.[375] Moreover, protecting databases in this way has been seriously considered in the past but not accepted in this country.[376] Although this type of system would not require an explicit allocation of government resources, users still bear its costs, both in terms of expenses and limitations on access.[377] In addition, privatizing data in this way would limit the ability to build upon and improve the data, similar to the problems described earlier with providing trade secret protection for the data generated as a result of leveraged patents.[378]

Similarly, prizes, tax breaks, or other rewards might provide sufficient incentives to innovate in this area without many of the downsides of patents, but they also are not without costs.[379] For instance, potential users of an invention will often bear its costs regardless of whether a patent or a prize provides the incentive for its creation.[380] Notwithstanding the issue of cost, data-generating patents seem particularly ill-suited for non-patent rewards.[381] Data-generating patents, which may have preemptive effects and be leveraged to collect data in unforeseeable secondary markets, are likely to be quite difficult for the government to value accurately ex ante.

3. *Tailoring Solutions Based on Facts and Further Empirical Research*.—In sum, each potential solution to innovation and disclosure-related concerns has potentially significant limitations. Nonetheless, it seems clear that most of the ex ante solutions—such as limiting patent term or narrowing patent eligibility—suffer from large error

374 *See* Council Directive 96/9, 1996 O.J. (L 77) 20 (discussing the rights of database producers), http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31996L0009 [https://perma.cc/99LJ-AM6U].

375 *See supra* Section II.B.3.

376 *See, e.g.*, Jane C. Ginsburg, *Copyright, Common Law, and* Sui Generis *Protection of Databases in the United States and Abroad*, 66 U. CIN. L. REV. 151, 171 (1997).

377 *See, e.g.*, Daniel J. Hemel & Lisa Larrimore Ouellette, *Beyond the Patents–Prizes Debate*, 92 TEX. L. REV. 303, 312 (2013); Price II, *supra* note 89, at 467.

378 *See supra* notes 245–46 and accompanying text.

379 *See, e.g.*, Hemel & Ouellette, *supra* note 377, at 310–13; Benjamin N. Roin, *Intellectual Property Versus Prizes: Reframing the Debate*, 81 U. CHI. L. REV. 999 (2014). There are other doctrinal and regulatory mechanisms for dealing with the privacy- and health-related concerns with aggregating medical data, though here we focus on the specific costs potentially imposed by data-generating patents. *See* Evans, *supra* note 287, at 77–82.

380 *See* Hemel & Ouellette, *supra* note 377, at 349–50.

381 *See id.* at 333–35 (discussing the cost and benefits of ex ante solutions); Richard A. Posner, *Intellectual Property: The Law and Economics Approach*, 19 J. ECON. PERSP. 57, 59 (2005) ("Calculating the optimal reward is difficult . . . .").

costs, as well as potentially substantial downstream costs in channeling these inventions into trade secrecy. As such, we disfavor them, particularly for innovation-related concerns. Mandatory disclosure of information produced by a narrow class of easily identifiable data-generating inventions might suffer less from these concerns, but without more information on the potentially negative effects on innovation incentives and the feasibility of monitoring compliance, we disfavor such solutions.

As for the ex post solutions, the options of enhanced independent invention, reverse engineering, and experimental use defenses—as well as limited denials of injunctive relief perhaps coupled with mandatory disclosure in extreme cases—could be promising if courts can properly tailor these remedies. Agencies can also implement ex post solutions, especially in the area of disclosure. For instance, one possibility would be to require contribution to a confidential government repository of data for certain classes of data-generating inventions, such as medical diagnostic tests, with release only after sufficient agency study of the information and its economic effects.

Ultimately, whether any of these solutions can be adequately implemented will depend on the competency of legislatures, courts, and agencies. Such competency can arguably be improved with more empirical study on the economic effects of data-generating patents. Here, we have begun that process by recognizing the problem of data-generating patents, with an eye towards potential solutions and further study.

CONCLUSION

We have discussed data-generating patents as a unique illustration of patents and trade secrets acting as economic complements. Data-generating patents, which by design generate valuable data by their use, may provide the patentee market power not just over the invention, but also over data generated by it. Trade secret law affords further protection for the generated data, even where the underlying patent allows its holder to preempt use or collection of the data in unforeseeable markets. Although the use of patents and trade secrets in this way may sometimes result in increased efficiency or provide additional innovation incentives in areas where they may be sorely lacking, in other circumstances they may produce detrimental effects. We have set forth factors for identifying potentially problematic data-generating patents and offered suggestions to mitigate the potential harm to innovation and disclosure that may result from their use. We hope our discussion serves as a useful starting point for future research into this complex intersection of intellectual property law and big data.